

Fine-Tuning LLM

For sentiment analysis used to classify people with mental health statuses

0. Imports

```
In [ ]: import numpy as np
import pandas as pd
import unsloth
import os
from tqdm import tqdm
import bitsandbytes as bnb
import torch
import torch.nn as nn
import transformers
from datasets import Dataset
from peft import LoraConfig, PeftConfig
from trl import SFTTrainer, SFTConfig
from trl import setup_chat_format
from transformers import (AutoModelForCausalLM,
                          AutoTokenizer,
                          BitsAndBytesConfig,
                          TrainingArguments,
                          pipeline,
                          logging)
from sklearn.metrics import (accuracy_score,
                             classification_report,
                             confusion_matrix)
from sklearn.model_selection import train_test_split
```

🐍 Unsloth: Will patch your computer to enable 2x faster free finetuning.

Unsloth: Failed to patch Gemma3ForConditionalGeneration.

🐍 Unsloth Zoo will now patch everything to make training faster!

1. Loading the Model

```
In [ ]: from unsloth import FastLanguageModel

max_seq_length = 1024 # Choose any! We auto support RoPE Scaling internally!
dtype = None # None for auto detection. Float16 for Tesla T4, V100, Bfloat16
load_in_4bit = True # Use 4bit quantization to reduce memory usage. Can be F

model, tokenizer = FastLanguageModel.from_pretrained(
    model_name = "unsloth/Meta-Llama-3.1-8B-Instruct-bnb-4bit",
    max_seq_length = max_seq_length,
    dtype = dtype,
    load_in_4bit = load_in_4bit,
    token = HF_TOKEN, # use one if using gated models like meta-llama/Llama-
```

```
)

# Set model configuration parameters similar to reference code
model.config.use_cache = False
model.config.pretraining_tp = 1

# Ensure pad token is set correctly
if tokenizer.pad_token_id is None:
    tokenizer.pad_token_id = tokenizer.eos_token_id

==(=====)== Unsloth 2025.3.19: Fast Llama patching. Transformers: 4.51.3.
  \ \ / | NVIDIA GeForce RTX 4090. Num GPUs = 1. Max memory: 23.542 GB.
Platform: Linux.
0^0/ \_/ \ Torch: 2.6.0+cu124. CUDA: 8.9. CUDA Toolkit: 12.4. Triton: 3.
2.0
\ _____ / Bfloat16 = TRUE. FA [Xformers = 0.0.29.post3. FA2 = False]
  "-_____" Free license: http://github.com/unslothai/unsloth
Unsloth: Fast downloading is enabled - ignore downloading bars which are red
colored!
```

2. Setting up wandb to track training

```
In [ ]: import wandb

wb_token = WANDB_TOKEN

wandb.login(key=wb_token)
run = wandb.init(
    project='Fine-tune llama-3.1-8b-it on Sentiment Analysis Dataset',
    job_type="training",
    anonymous="allow"
)
```

```
wandb: Using wandb-core as the SDK backend. Please refer to https://wandb.m
e/wandb-core for more information.
wandb: WARNING If you're specifying your api key in code, ensure this code i
s not shared publicly.
wandb: WARNING Consider setting the WANDB_API_KEY environment variable, or r
unning `wandb login` from the command line.
wandb: Appending key for api.wandb.ai to your netrc file: /home/makers/.netrc
wandb: Currently logged in as: dionyichia (dionyichia-nanyang-technological-
university-singapore) to https://api.wandb.ai. Use `wandb login --relogin` t
o force relogin
```

Tracking run with wandb version 0.19.9

Run data is saved locally in /home/makers/dion-dsai/wandb/run-
20250417_163054-00ozrnc9

Syncing run **dutiful-wind-19** to [Weights & Biases \(docs\)](https://docs.wandb.ai/)

View project at [https://wandb.ai/dionyichia-nanyang-technological-university-](https://wandb.ai/dionyichia-nanyang-technological-university-singapore/Fine-tune%20llama-3.1-8b-)
[singapore/Fine-tune%20llama-3.1-8b-](https://wandb.ai/dionyichia-nanyang-technological-university-singapore/Fine-tune%20llama-3.1-8b-)

[it%20on%20Sentiment%20Analysis%20Dataset?](https://wandb.ai/dionyichia-nanyang-technological-university-singapore/Fine-tune%20llama-3.1-8b-it%20on%20Sentiment%20Analysis%20Dataset?)

[apiKey=73ba7c69e019dd2879de1bd311b39552ea00e93a](https://wandb.ai/dionyichia-nanyang-technological-university-singapore/Fine-tune%20llama-3.1-8b-it%20on%20Sentiment%20Analysis%20Dataset?apiKey=73ba7c69e019dd2879de1bd311b39552ea00e93a)

View run at <https://wandb.ai/dionychia-nanyang-technological-university-singapore/Fine-tune%20llama-3.1-8b->

[it%20on%20Sentiment%20Analysis%20Dataset/runs/0oozrnc9?](https://wandb.ai/dionychia-nanyang-technological-university-singapore/Fine-tune%20llama-3.1-8b-it%20on%20Sentiment%20Analysis%20Dataset/runs/0oozrnc9?apiKey=73ba7c69e019dd2879de1bd311b39552ea00e93a)

[apiKey=73ba7c69e019dd2879de1bd311b39552ea00e93a](https://wandb.ai/dionychia-nanyang-technological-university-singapore/Fine-tune%20llama-3.1-8b-it%20on%20Sentiment%20Analysis%20Dataset/runs/0oozrnc9?apiKey=73ba7c69e019dd2879de1bd311b39552ea00e93a)

Do NOT share these links with anyone. They can be used to claim your runs.

3. Loading Dataset

```
In [ ]: # Read the data
df = pd.read_csv("./balanced_dataset.csv", index_col="Unnamed: 0")
df.head()

# Rebalance classes - Check to ensure that all statuses have the same number
min_class_size = df["status"].value_counts().min()
df = df.groupby("status", group_keys=False).apply(lambda x: x.sample(n=min_c
df = df.reset_index(drop=True)

print("\nBalanced class distribution:\n", df["status"].value_counts())

# Create a stratified train-eval-test split
# First split into train and temp (which will become eval and test)
features = df.drop("status", axis=1) # All columns except status
target = df["status"]

X_train, X_temp, y_train, y_temp = train_test_split(
    features,
    target,
    test_size=0.2, # 20% for eval+test combined
    random_state=42,
    stratify=target # This ensures class balance is maintained
)

# Then split temp into eval and test (50% each, which gives us 10% of origin
X_eval, X_test, y_eval, y_test = train_test_split(
    X_temp,
    y_temp,
    test_size=0.5,
    random_state=42,
    stratify=y_temp # This maintains class balance again
)

# Combine features and targets back for prompt generation
train_df = X_train.copy()
train_df["status"] = y_train
train_df = train_df.reset_index(drop=True)

eval_df = X_eval.copy()
eval_df["status"] = y_eval
eval_df = eval_df.reset_index(drop=True)

test_df = X_test.copy()
test_df["status"] = y_test
test_df = test_df.reset_index(drop=True)
```

```
# Verify class distribution in each split
print("\nTrain set class distribution:")
print(train_df["status"].value_counts())

print("\nEval set class distribution:")
print(eval_df["status"].value_counts())

print("\nTest set class distribution:")
print(test_df["status"].value_counts())
```

Balanced class distribution:

```
status
Anxiety      3888
Depression   3888
Normal       3888
Name: count, dtype: int64
```

Train set class distribution:

```
status
Anxiety      3111
Normal       3110
Depression   3110
Name: count, dtype: int64
```

Eval set class distribution:

```
status
Depression   389
Normal       389
Anxiety      388
Name: count, dtype: int64
```

Test set class distribution:

```
status
Normal       389
Anxiety      389
Depression   389
Name: count, dtype: int64
```

```
/tmp/ipykernel_1422688/193280215.py:9: DeprecationWarning: DataFrameGroupBy.apply operated on the grouping columns. This behavior is deprecated, and in a future version of pandas the grouping columns will be excluded from the operation. Either pass `include_groups=False` to exclude the groupings or explicitly select the grouping columns after groupby to silence this warning.
  df = df.groupby("status", group_keys=False).apply(lambda x: x.sample(n=min_class_size, random_state=85))
```

4. Prepping data for training

```
In [6]: # Define the prompt generation functions
def generate_prompt(data_point):
    return f"""
        Classify the text into Normal, Depression, Anxiety, Bipolar, and
        text: {data_point["statement"]}
        label: {data_point["status"]}"""
```

```

def generate_test_prompt(data_point):
    return f"""
        Classify the text into Normal, Depression, Anxiety, Bipolar, and
text: {data_point["statement"]}
label: """.strip()

# Generate prompts for training and evaluation data
train_df.loc[:, 'text'] = train_df.apply(generate_prompt, axis=1)
eval_df.loc[:, 'text'] = eval_df.apply(generate_prompt, axis=1)

# For test data, save the true labels separately before generating prompts
y_true = test_df["status"].copy()
test_prompts = test_df.apply(generate_test_prompt, axis=1)
test_df = pd.DataFrame({"text": test_prompts})

# Convert to datasets
train_data = Dataset.from_pandas(train_df[["text"]]) # Convert to datasets
eval_data = Dataset.from_pandas(eval_df[["text"]])
test_data = Dataset.from_pandas(test_df)

print("\nFinal dataset shapes:")
print(f"Train: {train_data.shape}")
print(f"Eval: {eval_data.shape}")
print(f"Test: {test_data.shape}")

```

Final dataset shapes:

Train: (9331, 1)

Eval: (1166, 1)

Test: (1167, 1)

In [7]: `train_data['text'][3]`

Out[7]: 'Classify the text into Normal, Depression, Anxiety, Bipolar, and return the answer as the corresponding mental health disorder label.\ntext: i apologized to her, but i could tell she was still upset.\nlabel: Normal'

```

In [8]: def predict(test, model, tokenizer):
        y_pred = []
        categories = ["Normal", "Depression", "Anxiety", "Bipolar"]

        for i in tqdm(range(len(test))):
            prompt = test.iloc[i]["text"]
            pipe = pipeline(task="text-generation",
                            model=model,
                            tokenizer=tokenizer,
                            max_new_tokens=2,
                            temperature=0.1)

            result = pipe(prompt)
            answer = result[0]['generated_text'].split("label: ")[-1].strip()

            # Determine the predicted category
            for category in categories:
                if category.lower() in answer.lower():
                    y_pred.append(category)
                    break

```

```
        else:
            y_pred.append("none")

    return y_pred

y_pred = predict(test_df, model, tokenizer)
```

0%		0/1167	[00:00<?, ?it/s]	Device set to use cuda:0
0%		1/1167	[00:02<39:30, 2.03s/it]	Device set to use cuda:0
0%		2/1167	[00:02<17:45, 1.09it/s]	Device set to use cuda:0
0%		3/1167	[00:02<10:45, 1.80it/s]	Device set to use cuda:0
0%		4/1167	[00:02<07:25, 2.61it/s]	Device set to use cuda:0
0%		5/1167	[00:02<05:33, 3.49it/s]	Device set to use cuda:0
1%		6/1167	[00:02<04:26, 4.36it/s]	Device set to use cuda:0
1%		7/1167	[00:02<03:53, 4.97it/s]	Device set to use cuda:0
1%		8/1167	[00:02<03:22, 5.72it/s]	Device set to use cuda:0
1%		9/1167	[00:03<03:04, 6.26it/s]	Device set to use cuda:0
1%		10/1167	[00:03<02:50, 6.80it/s]	Device set to use cuda:0
1%		11/1167	[00:03<02:39, 7.23it/s]	Device set to use cuda:0
1%		12/1167	[00:03<02:34, 7.48it/s]	Device set to use cuda:0
1%		13/1167	[00:03<02:29, 7.74it/s]	Device set to use cuda:0
1%		14/1167	[00:03<02:25, 7.93it/s]	Device set to use cuda:0
1%		15/1167	[00:03<02:21, 8.14it/s]	Device set to use cuda:0
1%		16/1167	[00:03<02:32, 7.52it/s]	Device set to use cuda:0
1%		17/1167	[00:04<02:28, 7.74it/s]	Device set to use cuda:0
2%		18/1167	[00:04<02:25, 7.92it/s]	Device set to use cuda:0
2%		19/1167	[00:04<02:21, 8.12it/s]	Device set to use cuda:0
2%		20/1167	[00:04<02:18, 8.30it/s]	Device set to use cuda:0
2%		21/1167	[00:04<02:17, 8.36it/s]	Device set to use cuda:0
2%		22/1167	[00:04<02:17, 8.34it/s]	Device set to use cuda:0
2%		23/1167	[00:04<02:19, 8.23it/s]	Device set to use cuda:0
2%		24/1167	[00:04<02:20, 8.16it/s]	Device set to use cuda:0
2%		25/1167	[00:04<02:19, 8.21it/s]	Device set to use cuda:0
2%		26/1167	[00:05<02:18, 8.26it/s]	Device set to use cuda:0
2%		27/1167	[00:05<02:17, 8.29it/s]	Device set to use cuda:0
2%		28/1167	[00:05<02:17, 8.30it/s]	Device set to use cuda:0
2%		29/1167	[00:05<02:15, 8.38it/s]	Device set to use cuda:0
3%		30/1167	[00:05<02:14, 8.45it/s]	Device set to use cuda:0
3%		31/1167	[00:05<02:13, 8.53it/s]	Device set to use cuda:0
3%		32/1167	[00:05<02:16, 8.32it/s]	Device set to use cuda:0
3%		33/1167	[00:05<02:21, 7.99it/s]	Device set to use cuda:0
3%		34/1167	[00:06<02:18, 8.18it/s]	Device set to use cuda:0
3%		35/1167	[00:06<02:17, 8.21it/s]	Device set to use cuda:0
3%		36/1167	[00:06<02:16, 8.27it/s]	Device set to use cuda:0
3%		37/1167	[00:06<02:16, 8.31it/s]	Device set to use cuda:0
3%		38/1167	[00:06<02:16, 8.29it/s]	Device set to use cuda:0
3%		39/1167	[00:06<02:16, 8.27it/s]	Device set to use cuda:0
3%		40/1167	[00:06<02:14, 8.35it/s]	Device set to use cuda:0
4%		41/1167	[00:06<02:14, 8.37it/s]	Device set to use cuda:0
4%		42/1167	[00:07<02:14, 8.37it/s]	Device set to use cuda:0
4%		43/1167	[00:07<02:14, 8.37it/s]	Device set to use cuda:0
4%		44/1167	[00:07<02:13, 8.41it/s]	Device set to use cuda:0
4%		45/1167	[00:07<02:13, 8.38it/s]	Device set to use cuda:0
4%		46/1167	[00:07<02:13, 8.41it/s]	Device set to use cuda:0
4%		47/1167	[00:07<02:13, 8.41it/s]	Device set to use cuda:0
4%		48/1167	[00:07<02:12, 8.42it/s]	Device set to use cuda:0
4%		49/1167	[00:07<02:15, 8.27it/s]	Device set to use cuda:0
4%		50/1167	[00:07<02:15, 8.21it/s]	Device set to use cuda:0
4%		51/1167	[00:08<02:17, 8.10it/s]	Device set to use cuda:0
4%		52/1167	[00:08<02:16, 8.15it/s]	Device set to use cuda:0
5%		53/1167	[00:08<02:16, 8.16it/s]	Device set to use cuda:0
5%		54/1167	[00:08<02:16, 8.15it/s]	Device set to use cuda:0
5%		55/1167	[00:08<02:25, 7.63it/s]	Device set to use cuda:0

5%		56/1167	[00:08<02:20,	7.92it/s]	Device set to use cuda:0
5%		57/1167	[00:08<02:19,	7.95it/s]	Device set to use cuda:0
5%		58/1167	[00:09<02:24,	7.68it/s]	Device set to use cuda:0
5%		59/1167	[00:09<02:20,	7.88it/s]	Device set to use cuda:0
5%		60/1167	[00:09<02:16,	8.13it/s]	Device set to use cuda:0
5%		61/1167	[00:09<02:12,	8.33it/s]	Device set to use cuda:0
5%		62/1167	[00:09<02:11,	8.38it/s]	Device set to use cuda:0
5%		63/1167	[00:09<02:13,	8.29it/s]	Device set to use cuda:0
5%		64/1167	[00:09<02:12,	8.34it/s]	Device set to use cuda:0
6%		65/1167	[00:09<02:16,	8.10it/s]	Device set to use cuda:0
6%		66/1167	[00:09<02:13,	8.24it/s]	Device set to use cuda:0
6%		67/1167	[00:10<02:11,	8.36it/s]	Device set to use cuda:0
6%		68/1167	[00:10<02:10,	8.44it/s]	Device set to use cuda:0
6%		69/1167	[00:10<02:10,	8.41it/s]	Device set to use cuda:0
6%		70/1167	[00:10<02:10,	8.40it/s]	Device set to use cuda:0
6%		71/1167	[00:10<02:08,	8.52it/s]	Device set to use cuda:0
6%		72/1167	[00:10<02:10,	8.39it/s]	Device set to use cuda:0
6%		73/1167	[00:10<02:11,	8.35it/s]	Device set to use cuda:0
6%		74/1167	[00:10<02:11,	8.33it/s]	Device set to use cuda:0
6%		75/1167	[00:11<02:10,	8.39it/s]	Device set to use cuda:0
7%		76/1167	[00:11<02:11,	8.31it/s]	Device set to use cuda:0
7%		77/1167	[00:11<02:10,	8.34it/s]	Device set to use cuda:0
7%		78/1167	[00:11<02:11,	8.30it/s]	Device set to use cuda:0
7%		79/1167	[00:11<02:10,	8.33it/s]	Device set to use cuda:0
7%		80/1167	[00:11<02:11,	8.30it/s]	Device set to use cuda:0
7%		81/1167	[00:11<02:09,	8.36it/s]	Device set to use cuda:0
7%		82/1167	[00:11<02:10,	8.29it/s]	Device set to use cuda:0
7%		83/1167	[00:11<02:08,	8.41it/s]	Device set to use cuda:0
7%		84/1167	[00:12<02:07,	8.50it/s]	Device set to use cuda:0
7%		85/1167	[00:12<02:08,	8.44it/s]	Device set to use cuda:0
7%		86/1167	[00:12<02:09,	8.33it/s]	Device set to use cuda:0
7%		87/1167	[00:12<02:10,	8.28it/s]	Device set to use cuda:0
8%		88/1167	[00:12<02:07,	8.44it/s]	Device set to use cuda:0
8%		89/1167	[00:12<02:11,	8.21it/s]	Device set to use cuda:0
8%		90/1167	[00:12<02:19,	7.70it/s]	Device set to use cuda:0
8%		91/1167	[00:13<02:22,	7.54it/s]	Device set to use cuda:0
8%		92/1167	[00:13<02:24,	7.45it/s]	Device set to use cuda:0
8%		93/1167	[00:13<02:39,	6.75it/s]	Device set to use cuda:0
8%		94/1167	[00:13<02:38,	6.77it/s]	Device set to use cuda:0
8%		95/1167	[00:13<02:29,	7.19it/s]	Device set to use cuda:0
8%		96/1167	[00:13<02:21,	7.59it/s]	Device set to use cuda:0
8%		97/1167	[00:13<02:17,	7.78it/s]	Device set to use cuda:0
8%		98/1167	[00:13<02:23,	7.44it/s]	Device set to use cuda:0
8%		99/1167	[00:14<02:21,	7.57it/s]	Device set to use cuda:0
9%		100/1167	[00:14<02:17,	7.74it/s]	Device set to use cuda:0

Unsloth: Input IDs of length 2131 > the model's max sequence length of 1024.
We shall truncate it ourselves. It's imperative if you correct this issue first.

9%		101/1167	[00:14<02:34,	6.91it/s]	Device set to use cuda:0
9%		102/1167	[00:14<02:27,	7.21it/s]	Device set to use cuda:0
9%		103/1167	[00:14<02:23,	7.39it/s]	Device set to use cuda:0
9%		104/1167	[00:14<02:19,	7.63it/s]	Device set to use cuda:0
9%		105/1167	[00:14<02:19,	7.61it/s]	Device set to use cuda:0
9%		106/1167	[00:15<02:15,	7.85it/s]	Device set to use cuda:0
9%		107/1167	[00:15<02:10,	8.13it/s]	Device set to use cuda:0
9%		108/1167	[00:15<02:07,	8.29it/s]	Device set to use cuda:0

9%			109/1167	[00:15<02:05,	8.42it/s]	Device set to use cuda:0
9%			110/1167	[00:15<02:03,	8.53it/s]	Device set to use cuda:0
10%			111/1167	[00:15<02:02,	8.59it/s]	Device set to use cuda:0
10%			112/1167	[00:15<02:03,	8.57it/s]	Device set to use cuda:0
10%			113/1167	[00:15<02:02,	8.60it/s]	Device set to use cuda:0
10%			114/1167	[00:15<02:03,	8.52it/s]	Device set to use cuda:0
10%			115/1167	[00:16<02:04,	8.47it/s]	Device set to use cuda:0
10%			116/1167	[00:16<02:03,	8.48it/s]	Device set to use cuda:0
10%			117/1167	[00:16<02:03,	8.48it/s]	Device set to use cuda:0
10%			118/1167	[00:16<02:18,	7.58it/s]	Device set to use cuda:0
10%			119/1167	[00:16<02:12,	7.90it/s]	Device set to use cuda:0
10%			120/1167	[00:16<02:10,	8.00it/s]	Device set to use cuda:0
10%			121/1167	[00:16<02:09,	8.05it/s]	Device set to use cuda:0
10%			122/1167	[00:16<02:16,	7.63it/s]	Device set to use cuda:0
11%			123/1167	[00:17<02:10,	7.98it/s]	Device set to use cuda:0
11%			124/1167	[00:17<02:07,	8.19it/s]	Device set to use cuda:0
11%			125/1167	[00:17<02:04,	8.37it/s]	Device set to use cuda:0
11%			126/1167	[00:17<02:03,	8.46it/s]	Device set to use cuda:0
11%			127/1167	[00:17<02:02,	8.50it/s]	Device set to use cuda:0
11%			128/1167	[00:17<02:02,	8.51it/s]	Device set to use cuda:0
11%			129/1167	[00:17<02:01,	8.55it/s]	Device set to use cuda:0
11%			130/1167	[00:17<02:19,	7.45it/s]	Device set to use cuda:0
11%			131/1167	[00:18<02:13,	7.75it/s]	Device set to use cuda:0
11%			132/1167	[00:18<02:20,	7.38it/s]	Device set to use cuda:0
11%			133/1167	[00:18<02:14,	7.68it/s]	Device set to use cuda:0
11%			134/1167	[00:18<02:17,	7.51it/s]	Device set to use cuda:0
12%			135/1167	[00:18<02:11,	7.84it/s]	Device set to use cuda:0
12%			136/1167	[00:18<02:14,	7.67it/s]	Device set to use cuda:0
12%			137/1167	[00:18<02:11,	7.82it/s]	Device set to use cuda:0
12%			138/1167	[00:18<02:08,	7.99it/s]	Device set to use cuda:0
12%			139/1167	[00:19<02:05,	8.17it/s]	Device set to use cuda:0
12%			140/1167	[00:19<02:03,	8.34it/s]	Device set to use cuda:0
12%			141/1167	[00:19<02:01,	8.48it/s]	Device set to use cuda:0
12%			142/1167	[00:19<02:00,	8.49it/s]	Device set to use cuda:0
12%			143/1167	[00:19<02:00,	8.46it/s]	Device set to use cuda:0
12%			144/1167	[00:19<02:02,	8.33it/s]	Device set to use cuda:0
12%			145/1167	[00:19<02:01,	8.41it/s]	Device set to use cuda:0
13%			146/1167	[00:19<02:08,	7.97it/s]	Device set to use cuda:0
13%			147/1167	[00:20<02:08,	7.94it/s]	Device set to use cuda:0
13%			148/1167	[00:20<02:05,	8.12it/s]	Device set to use cuda:0
13%			149/1167	[00:20<02:02,	8.30it/s]	Device set to use cuda:0
13%			150/1167	[00:20<02:20,	7.26it/s]	Device set to use cuda:0
13%			151/1167	[00:20<02:12,	7.65it/s]	Device set to use cuda:0
13%			152/1167	[00:20<02:08,	7.88it/s]	Device set to use cuda:0
13%			153/1167	[00:20<02:08,	7.89it/s]	Device set to use cuda:0
13%			154/1167	[00:20<02:13,	7.57it/s]	Device set to use cuda:0
13%			155/1167	[00:21<02:12,	7.62it/s]	Device set to use cuda:0
13%			156/1167	[00:21<02:14,	7.49it/s]	Device set to use cuda:0
13%			157/1167	[00:21<02:15,	7.47it/s]	Device set to use cuda:0
14%			158/1167	[00:21<02:17,	7.34it/s]	Device set to use cuda:0
14%			159/1167	[00:21<02:25,	6.94it/s]	Device set to use cuda:0
14%			160/1167	[00:21<02:20,	7.17it/s]	Device set to use cuda:0
14%			161/1167	[00:21<02:19,	7.23it/s]	Device set to use cuda:0
14%			162/1167	[00:22<02:22,	7.07it/s]	Device set to use cuda:0
14%			163/1167	[00:22<02:17,	7.30it/s]	Device set to use cuda:0
14%			164/1167	[00:22<02:18,	7.22it/s]	Device set to use cuda:0

14%	█		165/1167	[00:22<02:18,	7.25it/s]	Device set to use cuda:0
14%	█		166/1167	[00:22<02:14,	7.45it/s]	Device set to use cuda:0
14%	█		167/1167	[00:22<02:14,	7.41it/s]	Device set to use cuda:0
14%	█		168/1167	[00:22<02:11,	7.58it/s]	Device set to use cuda:0
14%	█		169/1167	[00:23<02:13,	7.50it/s]	Device set to use cuda:0
15%	█		170/1167	[00:23<02:13,	7.47it/s]	Device set to use cuda:0
15%	█		171/1167	[00:23<02:10,	7.65it/s]	Device set to use cuda:0
15%	█		172/1167	[00:23<02:11,	7.57it/s]	Device set to use cuda:0
15%	█		173/1167	[00:23<02:11,	7.57it/s]	Device set to use cuda:0
15%	█		174/1167	[00:23<02:06,	7.86it/s]	Device set to use cuda:0
15%	█		175/1167	[00:23<02:08,	7.72it/s]	Device set to use cuda:0
15%	█		176/1167	[00:23<02:06,	7.81it/s]	Device set to use cuda:0
15%	█		177/1167	[00:24<02:08,	7.69it/s]	Device set to use cuda:0
15%	█		178/1167	[00:24<02:10,	7.61it/s]	Device set to use cuda:0
15%	█		179/1167	[00:24<02:05,	7.85it/s]	Device set to use cuda:0
15%	█		180/1167	[00:24<02:08,	7.71it/s]	Device set to use cuda:0
16%	█		181/1167	[00:24<02:06,	7.76it/s]	Device set to use cuda:0
16%	█		182/1167	[00:24<02:07,	7.72it/s]	Device set to use cuda:0
16%	█		183/1167	[00:24<02:10,	7.57it/s]	Device set to use cuda:0
16%	█		184/1167	[00:24<02:09,	7.56it/s]	Device set to use cuda:0
16%	█		185/1167	[00:25<02:16,	7.21it/s]	Device set to use cuda:0
16%	█		186/1167	[00:25<02:13,	7.36it/s]	Device set to use cuda:0
16%	█		187/1167	[00:25<02:15,	7.22it/s]	Device set to use cuda:0
16%	█		188/1167	[00:25<02:10,	7.51it/s]	Device set to use cuda:0
16%	█		189/1167	[00:25<02:13,	7.34it/s]	Device set to use cuda:0
16%	█		190/1167	[00:25<02:11,	7.45it/s]	Device set to use cuda:0
16%	█		191/1167	[00:25<02:10,	7.46it/s]	Device set to use cuda:0
16%	█		192/1167	[00:26<02:10,	7.45it/s]	Device set to use cuda:0
17%	█		193/1167	[00:26<02:07,	7.66it/s]	Device set to use cuda:0
17%	█		194/1167	[00:26<02:09,	7.50it/s]	Device set to use cuda:0
17%	█		195/1167	[00:26<02:15,	7.17it/s]	Device set to use cuda:0
17%	█		196/1167	[00:26<02:18,	6.99it/s]	Device set to use cuda:0
17%	█		197/1167	[00:26<02:15,	7.14it/s]	Device set to use cuda:0
17%	█		198/1167	[00:26<02:15,	7.14it/s]	Device set to use cuda:0
17%	█		199/1167	[00:27<02:14,	7.21it/s]	Device set to use cuda:0
17%	█		200/1167	[00:27<02:14,	7.21it/s]	Device set to use cuda:0
17%	█		201/1167	[00:27<02:10,	7.42it/s]	Device set to use cuda:0
17%	█		202/1167	[00:27<02:11,	7.36it/s]	Device set to use cuda:0
17%	█		203/1167	[00:27<02:07,	7.57it/s]	Device set to use cuda:0
17%	█		204/1167	[00:27<02:04,	7.72it/s]	Device set to use cuda:0
18%	█		205/1167	[00:27<02:02,	7.84it/s]	Device set to use cuda:0
18%	█		206/1167	[00:27<02:02,	7.83it/s]	Device set to use cuda:0
18%	█		207/1167	[00:28<02:04,	7.72it/s]	Device set to use cuda:0
18%	█		208/1167	[00:28<02:06,	7.61it/s]	Device set to use cuda:0
18%	█		209/1167	[00:28<02:04,	7.71it/s]	Device set to use cuda:0
18%	█		210/1167	[00:28<02:00,	7.92it/s]	Device set to use cuda:0
18%	█		211/1167	[00:28<01:58,	8.03it/s]	Device set to use cuda:0
18%	█		212/1167	[00:28<01:59,	8.02it/s]	Device set to use cuda:0
18%	█		213/1167	[00:28<01:56,	8.17it/s]	Device set to use cuda:0
18%	█		214/1167	[00:28<01:56,	8.16it/s]	Device set to use cuda:0
18%	█		215/1167	[00:29<02:04,	7.67it/s]	Device set to use cuda:0
19%	█		216/1167	[00:29<02:02,	7.77it/s]	Device set to use cuda:0
19%	█		217/1167	[00:29<02:02,	7.78it/s]	Device set to use cuda:0
19%	█		218/1167	[00:29<02:01,	7.84it/s]	Device set to use cuda:0
19%	█		219/1167	[00:29<01:57,	8.05it/s]	Device set to use cuda:0
19%	█		220/1167	[00:29<01:56,	8.15it/s]	Device set to use cuda:0

19%	221/1167	[00:29<01:56,	8.13it/s]	Device set to use cuda:0
19%	222/1167	[00:29<02:03,	7.65it/s]	Device set to use cuda:0
19%	223/1167	[00:30<02:00,	7.81it/s]	Device set to use cuda:0
19%	224/1167	[00:30<02:08,	7.36it/s]	Device set to use cuda:0
19%	225/1167	[00:30<02:03,	7.64it/s]	Device set to use cuda:0
19%	226/1167	[00:30<01:59,	7.85it/s]	Device set to use cuda:0
19%	227/1167	[00:30<01:57,	7.97it/s]	Device set to use cuda:0
20%	228/1167	[00:30<01:58,	7.92it/s]	Device set to use cuda:0
20%	229/1167	[00:30<01:57,	7.96it/s]	Device set to use cuda:0
20%	230/1167	[00:30<02:01,	7.71it/s]	Device set to use cuda:0
20%	231/1167	[00:31<02:02,	7.62it/s]	Device set to use cuda:0
20%	232/1167	[00:31<01:58,	7.88it/s]	Device set to use cuda:0
20%	233/1167	[00:31<01:55,	8.11it/s]	Device set to use cuda:0
20%	234/1167	[00:31<01:55,	8.09it/s]	Device set to use cuda:0
20%	235/1167	[00:31<01:53,	8.23it/s]	Device set to use cuda:0
20%	236/1167	[00:31<01:52,	8.30it/s]	Device set to use cuda:0
20%	237/1167	[00:31<01:51,	8.33it/s]	Device set to use cuda:0
20%	238/1167	[00:31<01:50,	8.37it/s]	Device set to use cuda:0
20%	239/1167	[00:32<01:51,	8.32it/s]	Device set to use cuda:0
21%	240/1167	[00:32<01:50,	8.41it/s]	Device set to use cuda:0
21%	241/1167	[00:32<01:54,	8.08it/s]	Device set to use cuda:0
21%	242/1167	[00:32<01:53,	8.18it/s]	Device set to use cuda:0
21%	243/1167	[00:32<01:51,	8.27it/s]	Device set to use cuda:0
21%	244/1167	[00:32<01:51,	8.27it/s]	Device set to use cuda:0
21%	245/1167	[00:32<01:51,	8.29it/s]	Device set to use cuda:0
21%	246/1167	[00:32<01:50,	8.31it/s]	Device set to use cuda:0
21%	247/1167	[00:33<01:50,	8.32it/s]	Device set to use cuda:0
21%	248/1167	[00:33<01:50,	8.30it/s]	Device set to use cuda:0
21%	249/1167	[00:33<01:49,	8.37it/s]	Device set to use cuda:0
21%	250/1167	[00:33<01:49,	8.36it/s]	Device set to use cuda:0
22%	251/1167	[00:33<01:49,	8.39it/s]	Device set to use cuda:0
22%	252/1167	[00:33<01:48,	8.40it/s]	Device set to use cuda:0
22%	253/1167	[00:33<01:48,	8.44it/s]	Device set to use cuda:0
22%	254/1167	[00:33<01:47,	8.46it/s]	Device set to use cuda:0
22%	255/1167	[00:34<01:56,	7.81it/s]	Device set to use cuda:0
22%	256/1167	[00:34<01:53,	8.02it/s]	Device set to use cuda:0
22%	257/1167	[00:34<01:50,	8.20it/s]	Device set to use cuda:0
22%	258/1167	[00:34<01:49,	8.27it/s]	Device set to use cuda:0
22%	259/1167	[00:34<01:48,	8.35it/s]	Device set to use cuda:0
22%	260/1167	[00:34<01:48,	8.39it/s]	Device set to use cuda:0
22%	261/1167	[00:34<01:50,	8.23it/s]	Device set to use cuda:0
22%	262/1167	[00:34<01:49,	8.23it/s]	Device set to use cuda:0
23%	263/1167	[00:34<01:50,	8.20it/s]	Device set to use cuda:0
23%	264/1167	[00:35<01:49,	8.28it/s]	Device set to use cuda:0
23%	265/1167	[00:35<01:52,	8.03it/s]	Device set to use cuda:0
23%	266/1167	[00:35<01:52,	8.01it/s]	Device set to use cuda:0
23%	267/1167	[00:35<01:50,	8.11it/s]	Device set to use cuda:0
23%	268/1167	[00:35<01:49,	8.24it/s]	Device set to use cuda:0
23%	269/1167	[00:35<01:48,	8.27it/s]	Device set to use cuda:0
23%	270/1167	[00:35<01:49,	8.23it/s]	Device set to use cuda:0
23%	271/1167	[00:35<01:48,	8.27it/s]	Device set to use cuda:0
23%	272/1167	[00:36<01:49,	8.19it/s]	Device set to use cuda:0
23%	273/1167	[00:36<01:47,	8.30it/s]	Device set to use cuda:0
23%	274/1167	[00:36<01:47,	8.32it/s]	Device set to use cuda:0
24%	275/1167	[00:36<01:46,	8.41it/s]	Device set to use cuda:0
24%	276/1167	[00:36<01:45,	8.43it/s]	Device set to use cuda:0

24%	277/1167	[00:36<01:46,	8.34it/s]	Device set to use cuda:0
24%	278/1167	[00:36<01:46,	8.33it/s]	Device set to use cuda:0
24%	279/1167	[00:36<01:46,	8.31it/s]	Device set to use cuda:0
24%	280/1167	[00:37<01:47,	8.22it/s]	Device set to use cuda:0
24%	281/1167	[00:37<01:46,	8.32it/s]	Device set to use cuda:0
24%	282/1167	[00:37<01:54,	7.75it/s]	Device set to use cuda:0
24%	283/1167	[00:37<01:51,	7.93it/s]	Device set to use cuda:0
24%	284/1167	[00:37<01:49,	8.05it/s]	Device set to use cuda:0
24%	285/1167	[00:37<01:47,	8.19it/s]	Device set to use cuda:0
25%	286/1167	[00:37<01:46,	8.27it/s]	Device set to use cuda:0
25%	287/1167	[00:37<01:46,	8.27it/s]	Device set to use cuda:0
25%	288/1167	[00:38<01:53,	7.75it/s]	Device set to use cuda:0
25%	289/1167	[00:38<01:48,	8.06it/s]	Device set to use cuda:0
25%	290/1167	[00:38<01:47,	8.19it/s]	Device set to use cuda:0
25%	291/1167	[00:38<01:45,	8.27it/s]	Device set to use cuda:0
25%	292/1167	[00:38<01:46,	8.20it/s]	Device set to use cuda:0
25%	293/1167	[00:38<01:47,	8.11it/s]	Device set to use cuda:0
25%	294/1167	[00:38<01:47,	8.12it/s]	Device set to use cuda:0
25%	295/1167	[00:38<01:46,	8.16it/s]	Device set to use cuda:0
25%	296/1167	[00:39<01:47,	8.09it/s]	Device set to use cuda:0
25%	297/1167	[00:39<01:46,	8.14it/s]	Device set to use cuda:0
26%	298/1167	[00:39<01:45,	8.25it/s]	Device set to use cuda:0
26%	299/1167	[00:39<01:45,	8.21it/s]	Device set to use cuda:0
26%	300/1167	[00:39<01:44,	8.31it/s]	Device set to use cuda:0
26%	301/1167	[00:39<01:42,	8.42it/s]	Device set to use cuda:0
26%	302/1167	[00:39<01:43,	8.34it/s]	Device set to use cuda:0
26%	303/1167	[00:39<01:42,	8.40it/s]	Device set to use cuda:0
26%	304/1167	[00:39<01:42,	8.44it/s]	Device set to use cuda:0
26%	305/1167	[00:40<01:44,	8.25it/s]	Device set to use cuda:0
26%	306/1167	[00:40<01:42,	8.38it/s]	Device set to use cuda:0
26%	307/1167	[00:40<01:42,	8.40it/s]	Device set to use cuda:0
26%	308/1167	[00:40<01:42,	8.41it/s]	Device set to use cuda:0
26%	309/1167	[00:40<01:41,	8.46it/s]	Device set to use cuda:0
27%	310/1167	[00:40<01:41,	8.42it/s]	Device set to use cuda:0
27%	311/1167	[00:40<01:41,	8.47it/s]	Device set to use cuda:0
27%	312/1167	[00:40<01:41,	8.44it/s]	Device set to use cuda:0
27%	313/1167	[00:41<01:45,	8.06it/s]	Device set to use cuda:0
27%	314/1167	[00:41<01:46,	8.03it/s]	Device set to use cuda:0
27%	315/1167	[00:41<01:44,	8.12it/s]	Device set to use cuda:0
27%	316/1167	[00:41<01:44,	8.12it/s]	Device set to use cuda:0
27%	317/1167	[00:41<01:44,	8.16it/s]	Device set to use cuda:0
27%	318/1167	[00:41<01:43,	8.17it/s]	Device set to use cuda:0
27%	319/1167	[00:41<01:43,	8.19it/s]	Device set to use cuda:0
27%	320/1167	[00:41<01:43,	8.15it/s]	Device set to use cuda:0
28%	321/1167	[00:42<01:44,	8.10it/s]	Device set to use cuda:0
28%	322/1167	[00:42<01:50,	7.62it/s]	Device set to use cuda:0
28%	323/1167	[00:42<01:47,	7.87it/s]	Device set to use cuda:0
28%	324/1167	[00:42<01:45,	8.01it/s]	Device set to use cuda:0
28%	325/1167	[00:42<01:43,	8.10it/s]	Device set to use cuda:0
28%	326/1167	[00:42<01:41,	8.27it/s]	Device set to use cuda:0
28%	327/1167	[00:42<01:41,	8.28it/s]	Device set to use cuda:0
28%	328/1167	[00:42<01:42,	8.19it/s]	Device set to use cuda:0
28%	329/1167	[00:43<01:42,	8.15it/s]	Device set to use cuda:0
28%	330/1167	[00:43<01:42,	8.18it/s]	Device set to use cuda:0
28%	331/1167	[00:43<01:40,	8.33it/s]	Device set to use cuda:0
28%	332/1167	[00:43<01:39,	8.38it/s]	Device set to use cuda:0

29%	333/1167	[00:43<01:38,	8.48it/s]	Device set to use cuda:0
29%	334/1167	[00:43<01:38,	8.50it/s]	Device set to use cuda:0
29%	335/1167	[00:43<01:37,	8.51it/s]	Device set to use cuda:0
29%	336/1167	[00:43<01:38,	8.46it/s]	Device set to use cuda:0
29%	337/1167	[00:43<01:38,	8.41it/s]	Device set to use cuda:0
29%	338/1167	[00:44<01:39,	8.34it/s]	Device set to use cuda:0
29%	339/1167	[00:44<01:40,	8.26it/s]	Device set to use cuda:0
29%	340/1167	[00:44<01:39,	8.29it/s]	Device set to use cuda:0
29%	341/1167	[00:44<01:39,	8.29it/s]	Device set to use cuda:0
29%	342/1167	[00:44<01:39,	8.28it/s]	Device set to use cuda:0
29%	343/1167	[00:44<01:39,	8.28it/s]	Device set to use cuda:0
29%	344/1167	[00:44<01:38,	8.33it/s]	Device set to use cuda:0
30%	345/1167	[00:44<01:41,	8.13it/s]	Device set to use cuda:0
30%	346/1167	[00:45<01:42,	7.98it/s]	Device set to use cuda:0
30%	347/1167	[00:45<01:42,	8.03it/s]	Device set to use cuda:0
30%	348/1167	[00:45<01:40,	8.14it/s]	Device set to use cuda:0
30%	349/1167	[00:45<01:39,	8.20it/s]	Device set to use cuda:0
30%	350/1167	[00:45<01:39,	8.19it/s]	Device set to use cuda:0
30%	351/1167	[00:45<01:39,	8.23it/s]	Device set to use cuda:0
30%	352/1167	[00:45<01:38,	8.30it/s]	Device set to use cuda:0
30%	353/1167	[00:45<01:37,	8.36it/s]	Device set to use cuda:0
30%	354/1167	[00:46<01:45,	7.71it/s]	Device set to use cuda:0
30%	355/1167	[00:46<01:48,	7.49it/s]	Device set to use cuda:0
31%	356/1167	[00:46<01:44,	7.78it/s]	Device set to use cuda:0
31%	357/1167	[00:46<01:42,	7.92it/s]	Device set to use cuda:0
31%	358/1167	[00:46<01:40,	8.04it/s]	Device set to use cuda:0
31%	359/1167	[00:46<01:40,	8.07it/s]	Device set to use cuda:0
31%	360/1167	[00:46<01:38,	8.21it/s]	Device set to use cuda:0
31%	361/1167	[00:46<01:39,	8.11it/s]	Device set to use cuda:0
31%	362/1167	[00:47<01:38,	8.16it/s]	Device set to use cuda:0
31%	363/1167	[00:47<01:37,	8.25it/s]	Device set to use cuda:0
31%	364/1167	[00:47<01:36,	8.33it/s]	Device set to use cuda:0
31%	365/1167	[00:47<01:36,	8.34it/s]	Device set to use cuda:0
31%	366/1167	[00:47<01:36,	8.30it/s]	Device set to use cuda:0
31%	367/1167	[00:47<01:35,	8.41it/s]	Device set to use cuda:0
32%	368/1167	[00:47<01:34,	8.46it/s]	Device set to use cuda:0
32%	369/1167	[00:47<01:34,	8.48it/s]	Device set to use cuda:0
32%	370/1167	[00:48<01:34,	8.40it/s]	Device set to use cuda:0
32%	371/1167	[00:48<01:37,	8.15it/s]	Device set to use cuda:0
32%	372/1167	[00:48<01:36,	8.28it/s]	Device set to use cuda:0
32%	373/1167	[00:48<01:34,	8.37it/s]	Device set to use cuda:0
32%	374/1167	[00:48<01:33,	8.44it/s]	Device set to use cuda:0
32%	375/1167	[00:48<01:33,	8.50it/s]	Device set to use cuda:0
32%	376/1167	[00:48<01:39,	7.92it/s]	Device set to use cuda:0
32%	377/1167	[00:48<01:38,	8.06it/s]	Device set to use cuda:0
32%	378/1167	[00:49<01:38,	8.03it/s]	Device set to use cuda:0
32%	379/1167	[00:49<01:36,	8.14it/s]	Device set to use cuda:0
33%	380/1167	[00:49<01:35,	8.25it/s]	Device set to use cuda:0
33%	381/1167	[00:49<01:34,	8.30it/s]	Device set to use cuda:0
33%	382/1167	[00:49<01:41,	7.72it/s]	Device set to use cuda:0
33%	383/1167	[00:49<01:43,	7.58it/s]	Device set to use cuda:0
33%	384/1167	[00:49<01:40,	7.78it/s]	Device set to use cuda:0
33%	385/1167	[00:49<01:37,	7.99it/s]	Device set to use cuda:0
33%	386/1167	[00:50<01:36,	8.13it/s]	Device set to use cuda:0
33%	387/1167	[00:50<01:38,	7.90it/s]	Device set to use cuda:0
33%	388/1167	[00:50<01:35,	8.12it/s]	Device set to use cuda:0

33%		389/1167	[00:50<01:36,	8.06it/s]	Device set to use cuda:0
33%		390/1167	[00:50<01:34,	8.21it/s]	Device set to use cuda:0
34%		391/1167	[00:50<01:33,	8.29it/s]	Device set to use cuda:0
34%		392/1167	[00:50<01:34,	8.24it/s]	Device set to use cuda:0
34%		393/1167	[00:50<01:32,	8.34it/s]	Device set to use cuda:0
34%		394/1167	[00:50<01:33,	8.27it/s]	Device set to use cuda:0
34%		395/1167	[00:51<01:33,	8.24it/s]	Device set to use cuda:0
34%		396/1167	[00:51<01:32,	8.35it/s]	Device set to use cuda:0
34%		397/1167	[00:51<01:32,	8.37it/s]	Device set to use cuda:0
34%		398/1167	[00:51<01:30,	8.47it/s]	Device set to use cuda:0
34%		399/1167	[00:51<01:30,	8.47it/s]	Device set to use cuda:0
34%		400/1167	[00:51<01:30,	8.50it/s]	Device set to use cuda:0
34%		401/1167	[00:51<01:31,	8.41it/s]	Device set to use cuda:0
34%		402/1167	[00:51<01:32,	8.31it/s]	Device set to use cuda:0
35%		403/1167	[00:52<01:38,	7.80it/s]	Device set to use cuda:0
35%		404/1167	[00:52<01:40,	7.63it/s]	Device set to use cuda:0
35%		405/1167	[00:52<01:36,	7.93it/s]	Device set to use cuda:0
35%		406/1167	[00:52<01:33,	8.11it/s]	Device set to use cuda:0
35%		407/1167	[00:52<01:33,	8.15it/s]	Device set to use cuda:0
35%		408/1167	[00:52<01:31,	8.31it/s]	Device set to use cuda:0
35%		409/1167	[00:52<01:30,	8.38it/s]	Device set to use cuda:0
35%		410/1167	[00:52<01:29,	8.42it/s]	Device set to use cuda:0
35%		411/1167	[00:53<01:37,	7.72it/s]	Device set to use cuda:0
35%		412/1167	[00:53<01:34,	7.97it/s]	Device set to use cuda:0
35%		413/1167	[00:53<01:32,	8.12it/s]	Device set to use cuda:0
35%		414/1167	[00:53<01:31,	8.26it/s]	Device set to use cuda:0
36%		415/1167	[00:53<01:30,	8.31it/s]	Device set to use cuda:0
36%		416/1167	[00:53<01:30,	8.32it/s]	Device set to use cuda:0
36%		417/1167	[00:53<01:29,	8.37it/s]	Device set to use cuda:0
36%		418/1167	[00:53<01:32,	8.06it/s]	Device set to use cuda:0
36%		419/1167	[00:54<01:31,	8.20it/s]	Device set to use cuda:0
36%		420/1167	[00:54<01:31,	8.19it/s]	Device set to use cuda:0
36%		421/1167	[00:54<01:29,	8.33it/s]	Device set to use cuda:0
36%		422/1167	[00:54<01:28,	8.41it/s]	Device set to use cuda:0
36%		423/1167	[00:54<01:31,	8.09it/s]	Device set to use cuda:0
36%		424/1167	[00:54<01:31,	8.16it/s]	Device set to use cuda:0
36%		425/1167	[00:54<01:32,	8.04it/s]	Device set to use cuda:0
37%		426/1167	[00:54<01:30,	8.16it/s]	Device set to use cuda:0
37%		427/1167	[00:55<01:29,	8.26it/s]	Device set to use cuda:0
37%		428/1167	[00:55<01:32,	7.95it/s]	Device set to use cuda:0
37%		429/1167	[00:55<01:36,	7.62it/s]	Device set to use cuda:0
37%		430/1167	[00:55<01:33,	7.88it/s]	Device set to use cuda:0
37%		431/1167	[00:55<01:32,	7.96it/s]	Device set to use cuda:0
37%		432/1167	[00:55<01:30,	8.10it/s]	Device set to use cuda:0
37%		433/1167	[00:55<01:30,	8.15it/s]	Device set to use cuda:0
37%		434/1167	[00:55<01:29,	8.19it/s]	Device set to use cuda:0
37%		435/1167	[00:56<01:29,	8.22it/s]	Device set to use cuda:0
37%		436/1167	[00:56<01:31,	7.99it/s]	Device set to use cuda:0
37%		437/1167	[00:56<01:30,	8.11it/s]	Device set to use cuda:0
38%		438/1167	[00:56<01:28,	8.24it/s]	Device set to use cuda:0
38%		439/1167	[00:56<01:27,	8.32it/s]	Device set to use cuda:0
38%		440/1167	[00:56<01:27,	8.35it/s]	Device set to use cuda:0
38%		441/1167	[00:56<01:27,	8.31it/s]	Device set to use cuda:0
38%		442/1167	[00:56<01:27,	8.29it/s]	Device set to use cuda:0
38%		443/1167	[00:56<01:27,	8.29it/s]	Device set to use cuda:0
38%		444/1167	[00:57<01:31,	7.94it/s]	Device set to use cuda:0

38%		445/1167	[00:57<01:30,	8.00it/s]	Device set to use cuda:0
38%		446/1167	[00:57<01:30,	7.98it/s]	Device set to use cuda:0
38%		447/1167	[00:57<01:29,	8.03it/s]	Device set to use cuda:0
38%		448/1167	[00:57<01:28,	8.11it/s]	Device set to use cuda:0
38%		449/1167	[00:57<01:33,	7.65it/s]	Device set to use cuda:0
39%		450/1167	[00:57<01:31,	7.86it/s]	Device set to use cuda:0
39%		451/1167	[00:57<01:28,	8.06it/s]	Device set to use cuda:0
39%		452/1167	[00:58<01:29,	8.01it/s]	Device set to use cuda:0
39%		453/1167	[00:58<01:28,	8.07it/s]	Device set to use cuda:0
39%		454/1167	[00:58<01:27,	8.16it/s]	Device set to use cuda:0
39%		455/1167	[00:58<01:26,	8.21it/s]	Device set to use cuda:0
39%		456/1167	[00:58<01:27,	8.16it/s]	Device set to use cuda:0
39%		457/1167	[00:58<01:28,	8.03it/s]	Device set to use cuda:0
39%		458/1167	[00:58<01:26,	8.16it/s]	Device set to use cuda:0
39%		459/1167	[00:58<01:27,	8.08it/s]	Device set to use cuda:0
39%		460/1167	[00:59<01:27,	8.05it/s]	Device set to use cuda:0
40%		461/1167	[00:59<01:30,	7.77it/s]	Device set to use cuda:0
40%		462/1167	[00:59<01:28,	7.95it/s]	Device set to use cuda:0
40%		463/1167	[00:59<01:28,	7.93it/s]	Device set to use cuda:0
40%		464/1167	[00:59<01:29,	7.89it/s]	Device set to use cuda:0
40%		465/1167	[00:59<01:28,	7.95it/s]	Device set to use cuda:0
40%		466/1167	[00:59<01:28,	7.90it/s]	Device set to use cuda:0
40%		467/1167	[00:59<01:28,	7.92it/s]	Device set to use cuda:0
40%		468/1167	[01:00<01:29,	7.82it/s]	Device set to use cuda:0
40%		469/1167	[01:00<01:27,	7.99it/s]	Device set to use cuda:0
40%		470/1167	[01:00<01:26,	8.10it/s]	Device set to use cuda:0
40%		471/1167	[01:00<01:30,	7.68it/s]	Device set to use cuda:0
40%		472/1167	[01:00<01:28,	7.87it/s]	Device set to use cuda:0
41%		473/1167	[01:00<01:27,	7.93it/s]	Device set to use cuda:0
41%		474/1167	[01:00<01:26,	8.04it/s]	Device set to use cuda:0
41%		475/1167	[01:00<01:26,	8.02it/s]	Device set to use cuda:0
41%		476/1167	[01:01<01:26,	8.02it/s]	Device set to use cuda:0
41%		477/1167	[01:01<01:27,	7.85it/s]	Device set to use cuda:0
41%		478/1167	[01:01<01:27,	7.85it/s]	Device set to use cuda:0
41%		479/1167	[01:01<01:26,	7.94it/s]	Device set to use cuda:0
41%		480/1167	[01:01<01:25,	8.01it/s]	Device set to use cuda:0
41%		481/1167	[01:01<01:27,	7.82it/s]	Device set to use cuda:0
41%		482/1167	[01:01<01:25,	7.99it/s]	Device set to use cuda:0
41%		483/1167	[01:01<01:24,	8.06it/s]	Device set to use cuda:0
41%		484/1167	[01:02<01:25,	7.95it/s]	Device set to use cuda:0
42%		485/1167	[01:02<01:25,	7.98it/s]	Device set to use cuda:0
42%		486/1167	[01:02<01:23,	8.12it/s]	Device set to use cuda:0
42%		487/1167	[01:02<01:23,	8.15it/s]	Device set to use cuda:0
42%		488/1167	[01:02<01:22,	8.19it/s]	Device set to use cuda:0
42%		489/1167	[01:02<01:22,	8.21it/s]	Device set to use cuda:0
42%		490/1167	[01:02<01:34,	7.19it/s]	Device set to use cuda:0
42%		491/1167	[01:03<01:31,	7.41it/s]	Device set to use cuda:0
42%		492/1167	[01:03<01:29,	7.56it/s]	Device set to use cuda:0
42%		493/1167	[01:03<01:32,	7.26it/s]	Device set to use cuda:0
42%		494/1167	[01:03<01:29,	7.53it/s]	Device set to use cuda:0
42%		495/1167	[01:03<01:27,	7.69it/s]	Device set to use cuda:0
43%		496/1167	[01:03<01:25,	7.84it/s]	Device set to use cuda:0
43%		497/1167	[01:03<01:25,	7.82it/s]	Device set to use cuda:0
43%		498/1167	[01:03<01:24,	7.94it/s]	Device set to use cuda:0
43%		499/1167	[01:04<01:23,	7.97it/s]	Device set to use cuda:0
43%		500/1167	[01:04<01:23,	7.95it/s]	Device set to use cuda:0

43%		501/1167	[01:04<01:22,	8.08it/s]	Device set to use cuda:0
43%		502/1167	[01:04<01:21,	8.18it/s]	Device set to use cuda:0
43%		503/1167	[01:04<01:22,	8.03it/s]	Device set to use cuda:0
43%		504/1167	[01:04<01:22,	8.07it/s]	Device set to use cuda:0
43%		505/1167	[01:04<01:22,	7.98it/s]	Device set to use cuda:0
43%		506/1167	[01:04<01:23,	7.95it/s]	Device set to use cuda:0
43%		507/1167	[01:05<01:22,	8.00it/s]	Device set to use cuda:0
44%		508/1167	[01:05<01:21,	8.04it/s]	Device set to use cuda:0
44%		509/1167	[01:05<01:21,	8.05it/s]	Device set to use cuda:0
44%		510/1167	[01:05<01:22,	7.98it/s]	Device set to use cuda:0
44%		511/1167	[01:05<01:21,	8.05it/s]	Device set to use cuda:0
44%		512/1167	[01:05<01:20,	8.09it/s]	Device set to use cuda:0
44%		513/1167	[01:05<01:20,	8.10it/s]	Device set to use cuda:0
44%		514/1167	[01:05<01:20,	8.09it/s]	Device set to use cuda:0
44%		515/1167	[01:06<01:29,	7.24it/s]	Device set to use cuda:0
44%		516/1167	[01:06<01:29,	7.26it/s]	Device set to use cuda:0
44%		517/1167	[01:06<01:26,	7.49it/s]	Device set to use cuda:0
44%		518/1167	[01:06<01:24,	7.71it/s]	Device set to use cuda:0
44%		519/1167	[01:06<01:22,	7.88it/s]	Device set to use cuda:0
45%		520/1167	[01:06<01:20,	8.05it/s]	Device set to use cuda:0
45%		521/1167	[01:06<01:19,	8.10it/s]	Device set to use cuda:0
45%		522/1167	[01:06<01:19,	8.15it/s]	Device set to use cuda:0
45%		523/1167	[01:07<01:19,	8.11it/s]	Device set to use cuda:0
45%		524/1167	[01:07<01:19,	8.14it/s]	Device set to use cuda:0
45%		525/1167	[01:07<01:22,	7.74it/s]	Device set to use cuda:0
45%		526/1167	[01:07<01:20,	7.92it/s]	Device set to use cuda:0
45%		527/1167	[01:07<01:19,	8.04it/s]	Device set to use cuda:0
45%		528/1167	[01:07<01:18,	8.15it/s]	Device set to use cuda:0
45%		529/1167	[01:07<01:17,	8.22it/s]	Device set to use cuda:0
45%		530/1167	[01:07<01:17,	8.21it/s]	Device set to use cuda:0
46%		531/1167	[01:08<01:17,	8.23it/s]	Device set to use cuda:0
46%		532/1167	[01:08<01:17,	8.15it/s]	Device set to use cuda:0
46%		533/1167	[01:08<01:17,	8.20it/s]	Device set to use cuda:0
46%		534/1167	[01:08<01:16,	8.26it/s]	Device set to use cuda:0
46%		535/1167	[01:08<01:16,	8.27it/s]	Device set to use cuda:0
46%		536/1167	[01:08<01:22,	7.65it/s]	Device set to use cuda:0
46%		537/1167	[01:08<01:22,	7.67it/s]	Device set to use cuda:0
46%		538/1167	[01:08<01:21,	7.68it/s]	Device set to use cuda:0
46%		539/1167	[01:09<01:21,	7.74it/s]	Device set to use cuda:0
46%		540/1167	[01:09<01:19,	7.92it/s]	Device set to use cuda:0
46%		541/1167	[01:09<01:17,	8.07it/s]	Device set to use cuda:0
46%		542/1167	[01:09<01:17,	8.02it/s]	Device set to use cuda:0
47%		543/1167	[01:09<01:16,	8.10it/s]	Device set to use cuda:0
47%		544/1167	[01:09<01:16,	8.17it/s]	Device set to use cuda:0
47%		545/1167	[01:09<01:15,	8.22it/s]	Device set to use cuda:0
47%		546/1167	[01:09<01:15,	8.26it/s]	Device set to use cuda:0
47%		547/1167	[01:10<01:14,	8.31it/s]	Device set to use cuda:0
47%		548/1167	[01:10<01:17,	8.02it/s]	Device set to use cuda:0
47%		549/1167	[01:10<01:15,	8.15it/s]	Device set to use cuda:0
47%		550/1167	[01:10<01:14,	8.29it/s]	Device set to use cuda:0
47%		551/1167	[01:10<01:13,	8.34it/s]	Device set to use cuda:0
47%		552/1167	[01:10<01:13,	8.36it/s]	Device set to use cuda:0
47%		553/1167	[01:10<01:13,	8.33it/s]	Device set to use cuda:0
47%		554/1167	[01:10<01:13,	8.36it/s]	Device set to use cuda:0
48%		555/1167	[01:11<01:13,	8.33it/s]	Device set to use cuda:0
48%		556/1167	[01:11<01:13,	8.30it/s]	Device set to use cuda:0

48%		557/1167	[01:11<01:19,	7.72it/s]	Device set to use cuda:0
48%		558/1167	[01:11<01:19,	7.70it/s]	Device set to use cuda:0
48%		559/1167	[01:11<01:17,	7.84it/s]	Device set to use cuda:0
48%		560/1167	[01:11<01:16,	7.97it/s]	Device set to use cuda:0
48%		561/1167	[01:11<01:16,	7.89it/s]	Device set to use cuda:0
48%		562/1167	[01:11<01:15,	7.97it/s]	Device set to use cuda:0
48%		563/1167	[01:12<01:14,	8.09it/s]	Device set to use cuda:0
48%		564/1167	[01:12<01:15,	7.98it/s]	Device set to use cuda:0
48%		565/1167	[01:12<01:14,	8.06it/s]	Device set to use cuda:0
49%		566/1167	[01:12<01:13,	8.16it/s]	Device set to use cuda:0
49%		567/1167	[01:12<01:14,	8.08it/s]	Device set to use cuda:0
49%		568/1167	[01:12<01:14,	8.06it/s]	Device set to use cuda:0
49%		569/1167	[01:12<01:13,	8.08it/s]	Device set to use cuda:0
49%		570/1167	[01:12<01:12,	8.21it/s]	Device set to use cuda:0
49%		571/1167	[01:13<01:12,	8.22it/s]	Device set to use cuda:0
49%		572/1167	[01:13<01:14,	7.98it/s]	Device set to use cuda:0
49%		573/1167	[01:13<01:14,	8.00it/s]	Device set to use cuda:0
49%		574/1167	[01:13<01:13,	8.03it/s]	Device set to use cuda:0
49%		575/1167	[01:13<01:14,	7.92it/s]	Device set to use cuda:0
49%		576/1167	[01:13<01:14,	7.96it/s]	Device set to use cuda:0
49%		577/1167	[01:13<01:17,	7.59it/s]	Device set to use cuda:0
50%		578/1167	[01:13<01:16,	7.69it/s]	Device set to use cuda:0
50%		579/1167	[01:14<01:14,	7.91it/s]	Device set to use cuda:0
50%		580/1167	[01:14<01:21,	7.24it/s]	Device set to use cuda:0
50%		581/1167	[01:14<01:18,	7.45it/s]	Device set to use cuda:0
50%		582/1167	[01:14<01:17,	7.59it/s]	Device set to use cuda:0
50%		583/1167	[01:14<01:18,	7.43it/s]	Device set to use cuda:0
50%		584/1167	[01:14<01:17,	7.55it/s]	Device set to use cuda:0
50%		585/1167	[01:14<01:17,	7.55it/s]	Device set to use cuda:0
50%		586/1167	[01:14<01:16,	7.58it/s]	Device set to use cuda:0
50%		587/1167	[01:15<01:16,	7.62it/s]	Device set to use cuda:0
50%		588/1167	[01:15<01:15,	7.70it/s]	Device set to use cuda:0
50%		589/1167	[01:15<01:23,	6.95it/s]	Device set to use cuda:0
51%		590/1167	[01:15<01:24,	6.80it/s]	Device set to use cuda:0
51%		591/1167	[01:15<01:20,	7.14it/s]	Device set to use cuda:0
51%		592/1167	[01:15<01:18,	7.36it/s]	Device set to use cuda:0
51%		593/1167	[01:15<01:15,	7.58it/s]	Device set to use cuda:0
51%		594/1167	[01:16<01:13,	7.85it/s]	Device set to use cuda:0
51%		595/1167	[01:16<01:14,	7.63it/s]	Device set to use cuda:0
51%		596/1167	[01:16<01:13,	7.81it/s]	Device set to use cuda:0
51%		597/1167	[01:16<01:16,	7.41it/s]	Device set to use cuda:0
51%		598/1167	[01:16<01:14,	7.68it/s]	Device set to use cuda:0
51%		599/1167	[01:16<01:14,	7.66it/s]	Device set to use cuda:0
51%		600/1167	[01:16<01:13,	7.73it/s]	Device set to use cuda:0
51%		601/1167	[01:16<01:12,	7.85it/s]	Device set to use cuda:0
52%		602/1167	[01:17<01:11,	7.91it/s]	Device set to use cuda:0
52%		603/1167	[01:17<01:11,	7.90it/s]	Device set to use cuda:0
52%		604/1167	[01:17<01:10,	8.02it/s]	Device set to use cuda:0
52%		605/1167	[01:17<01:10,	8.02it/s]	Device set to use cuda:0
52%		606/1167	[01:17<01:10,	7.97it/s]	Device set to use cuda:0
52%		607/1167	[01:17<01:09,	8.08it/s]	Device set to use cuda:0
52%		608/1167	[01:17<01:08,	8.17it/s]	Device set to use cuda:0
52%		609/1167	[01:17<01:08,	8.17it/s]	Device set to use cuda:0
52%		610/1167	[01:18<01:07,	8.20it/s]	Device set to use cuda:0
52%		611/1167	[01:18<01:18,	7.09it/s]	Device set to use cuda:0
52%		612/1167	[01:18<01:15,	7.31it/s]	Device set to use cuda:0

53%		613/1167	[01:18<01:13,	7.57it/s]	Device set to use cuda:0
53%		614/1167	[01:18<01:11,	7.70it/s]	Device set to use cuda:0
53%		615/1167	[01:18<01:11,	7.71it/s]	Device set to use cuda:0
53%		616/1167	[01:18<01:09,	7.91it/s]	Device set to use cuda:0
53%		617/1167	[01:19<01:13,	7.44it/s]	Device set to use cuda:0
53%		618/1167	[01:19<01:12,	7.57it/s]	Device set to use cuda:0
53%		619/1167	[01:19<01:13,	7.50it/s]	Device set to use cuda:0
53%		620/1167	[01:19<01:10,	7.71it/s]	Device set to use cuda:0
53%		621/1167	[01:19<01:16,	7.10it/s]	Device set to use cuda:0
53%		622/1167	[01:19<01:13,	7.39it/s]	Device set to use cuda:0
53%		623/1167	[01:19<01:11,	7.61it/s]	Device set to use cuda:0
53%		624/1167	[01:19<01:09,	7.81it/s]	Device set to use cuda:0
54%		625/1167	[01:20<01:08,	7.97it/s]	Device set to use cuda:0
54%		626/1167	[01:20<01:08,	7.87it/s]	Device set to use cuda:0
54%		627/1167	[01:20<01:09,	7.75it/s]	Device set to use cuda:0
54%		628/1167	[01:20<01:08,	7.92it/s]	Device set to use cuda:0
54%		629/1167	[01:20<01:06,	8.05it/s]	Device set to use cuda:0
54%		630/1167	[01:20<01:06,	8.03it/s]	Device set to use cuda:0
54%		631/1167	[01:20<01:06,	8.02it/s]	Device set to use cuda:0
54%		632/1167	[01:20<01:06,	8.01it/s]	Device set to use cuda:0
54%		633/1167	[01:21<01:06,	8.02it/s]	Device set to use cuda:0
54%		634/1167	[01:21<01:07,	7.93it/s]	Device set to use cuda:0
54%		635/1167	[01:21<01:06,	7.99it/s]	Device set to use cuda:0
54%		636/1167	[01:21<01:09,	7.66it/s]	Device set to use cuda:0
55%		637/1167	[01:21<01:07,	7.80it/s]	Device set to use cuda:0
55%		638/1167	[01:21<01:06,	8.01it/s]	Device set to use cuda:0
55%		639/1167	[01:21<01:05,	8.06it/s]	Device set to use cuda:0
55%		640/1167	[01:21<01:04,	8.18it/s]	Device set to use cuda:0
55%		641/1167	[01:22<01:04,	8.17it/s]	Device set to use cuda:0
55%		642/1167	[01:22<01:04,	8.17it/s]	Device set to use cuda:0
55%		643/1167	[01:22<01:04,	8.09it/s]	Device set to use cuda:0
55%		644/1167	[01:22<01:04,	8.14it/s]	Device set to use cuda:0
55%		645/1167	[01:22<01:03,	8.18it/s]	Device set to use cuda:0
55%		646/1167	[01:22<01:03,	8.25it/s]	Device set to use cuda:0
55%		647/1167	[01:22<01:03,	8.18it/s]	Device set to use cuda:0
56%		648/1167	[01:22<01:05,	7.87it/s]	Device set to use cuda:0
56%		649/1167	[01:23<01:04,	8.04it/s]	Device set to use cuda:0
56%		650/1167	[01:23<01:03,	8.11it/s]	Device set to use cuda:0
56%		651/1167	[01:23<01:03,	8.18it/s]	Device set to use cuda:0
56%		652/1167	[01:23<01:02,	8.30it/s]	Device set to use cuda:0
56%		653/1167	[01:23<01:02,	8.21it/s]	Device set to use cuda:0
56%		654/1167	[01:23<01:04,	7.94it/s]	Device set to use cuda:0
56%		655/1167	[01:23<01:08,	7.52it/s]	Device set to use cuda:0
56%		656/1167	[01:24<01:14,	6.86it/s]	Device set to use cuda:0

Unsloth: Input IDs of length 1861 > the model's max sequence length of 1024.
We shall truncate it ourselves. It's imperative if you correct this issue first.

56%		657/1167	[01:24<01:19,	6.40it/s]	Device set to use cuda:0
56%		658/1167	[01:24<01:14,	6.83it/s]	Device set to use cuda:0
56%		659/1167	[01:24<01:10,	7.21it/s]	Device set to use cuda:0
57%		660/1167	[01:24<01:07,	7.51it/s]	Device set to use cuda:0
57%		661/1167	[01:24<01:05,	7.76it/s]	Device set to use cuda:0
57%		662/1167	[01:24<01:03,	7.90it/s]	Device set to use cuda:0
57%		663/1167	[01:24<01:02,	8.09it/s]	Device set to use cuda:0
57%		664/1167	[01:25<01:01,	8.15it/s]	Device set to use cuda:0
57%		665/1167	[01:25<01:00,	8.25it/s]	Device set to use cuda:0

57%		666/1167	[01:25<01:00,	8.22it/s]	Device set to use cuda:0
57%		667/1167	[01:25<01:00,	8.26it/s]	Device set to use cuda:0
57%		668/1167	[01:25<00:59,	8.34it/s]	Device set to use cuda:0
57%		669/1167	[01:25<00:59,	8.31it/s]	Device set to use cuda:0
57%		670/1167	[01:25<01:00,	8.22it/s]	Device set to use cuda:0
57%		671/1167	[01:25<01:00,	8.22it/s]	Device set to use cuda:0
58%		672/1167	[01:26<01:00,	8.20it/s]	Device set to use cuda:0
58%		673/1167	[01:26<01:03,	7.80it/s]	Device set to use cuda:0
58%		674/1167	[01:26<01:03,	7.80it/s]	Device set to use cuda:0
58%		675/1167	[01:26<01:03,	7.70it/s]	Device set to use cuda:0
58%		676/1167	[01:26<01:02,	7.90it/s]	Device set to use cuda:0
58%		677/1167	[01:26<01:02,	7.88it/s]	Device set to use cuda:0
58%		678/1167	[01:26<01:01,	7.90it/s]	Device set to use cuda:0
58%		679/1167	[01:26<01:01,	7.93it/s]	Device set to use cuda:0
58%		680/1167	[01:27<01:01,	7.94it/s]	Device set to use cuda:0
58%		681/1167	[01:27<01:00,	7.98it/s]	Device set to use cuda:0
58%		682/1167	[01:27<01:02,	7.82it/s]	Device set to use cuda:0
59%		683/1167	[01:27<01:00,	8.02it/s]	Device set to use cuda:0
59%		684/1167	[01:27<01:00,	8.03it/s]	Device set to use cuda:0
59%		685/1167	[01:27<00:59,	8.15it/s]	Device set to use cuda:0
59%		686/1167	[01:27<00:59,	8.04it/s]	Device set to use cuda:0
59%		687/1167	[01:27<00:59,	8.13it/s]	Device set to use cuda:0
59%		688/1167	[01:28<00:58,	8.23it/s]	Device set to use cuda:0
59%		689/1167	[01:28<00:58,	8.22it/s]	Device set to use cuda:0
59%		690/1167	[01:28<01:02,	7.67it/s]	Device set to use cuda:0
59%		691/1167	[01:28<01:03,	7.47it/s]	Device set to use cuda:0
59%		692/1167	[01:28<01:02,	7.66it/s]	Device set to use cuda:0
59%		693/1167	[01:28<01:00,	7.84it/s]	Device set to use cuda:0
59%		694/1167	[01:28<00:59,	7.89it/s]	Device set to use cuda:0
60%		695/1167	[01:28<00:59,	7.93it/s]	Device set to use cuda:0
60%		696/1167	[01:29<00:59,	7.96it/s]	Device set to use cuda:0
60%		697/1167	[01:29<00:58,	8.06it/s]	Device set to use cuda:0
60%		698/1167	[01:29<00:58,	8.05it/s]	Device set to use cuda:0
60%		699/1167	[01:29<00:57,	8.13it/s]	Device set to use cuda:0
60%		700/1167	[01:29<00:59,	7.88it/s]	Device set to use cuda:0
60%		701/1167	[01:29<01:00,	7.75it/s]	Device set to use cuda:0
60%		702/1167	[01:29<01:00,	7.74it/s]	Device set to use cuda:0
60%		703/1167	[01:29<00:59,	7.84it/s]	Device set to use cuda:0
60%		704/1167	[01:30<00:59,	7.85it/s]	Device set to use cuda:0
60%		705/1167	[01:30<00:58,	7.88it/s]	Device set to use cuda:0
60%		706/1167	[01:30<00:58,	7.82it/s]	Device set to use cuda:0
61%		707/1167	[01:30<00:58,	7.89it/s]	Device set to use cuda:0
61%		708/1167	[01:30<00:57,	7.94it/s]	Device set to use cuda:0
61%		709/1167	[01:30<00:56,	8.06it/s]	Device set to use cuda:0
61%		710/1167	[01:30<00:59,	7.72it/s]	Device set to use cuda:0
61%		711/1167	[01:30<00:58,	7.83it/s]	Device set to use cuda:0
61%		712/1167	[01:31<00:57,	7.91it/s]	Device set to use cuda:0
61%		713/1167	[01:31<00:59,	7.67it/s]	Device set to use cuda:0
61%		714/1167	[01:31<00:58,	7.74it/s]	Device set to use cuda:0
61%		715/1167	[01:31<00:57,	7.91it/s]	Device set to use cuda:0
61%		716/1167	[01:31<00:56,	7.92it/s]	Device set to use cuda:0
61%		717/1167	[01:31<00:56,	7.95it/s]	Device set to use cuda:0
62%		718/1167	[01:31<00:57,	7.83it/s]	Device set to use cuda:0
62%		719/1167	[01:31<00:56,	7.92it/s]	Device set to use cuda:0
62%		720/1167	[01:32<00:55,	8.01it/s]	Device set to use cuda:0
62%		721/1167	[01:32<00:56,	7.92it/s]	Device set to use cuda:0

62%		722/1167	[01:32<00:56,	7.92it/s]	Device set to use cuda:0
62%		723/1167	[01:32<00:55,	8.05it/s]	Device set to use cuda:0
62%		724/1167	[01:32<00:54,	8.14it/s]	Device set to use cuda:0
62%		725/1167	[01:32<00:54,	8.17it/s]	Device set to use cuda:0
62%		726/1167	[01:32<00:54,	8.12it/s]	Device set to use cuda:0
62%		727/1167	[01:32<00:54,	8.07it/s]	Device set to use cuda:0
62%		728/1167	[01:33<00:53,	8.19it/s]	Device set to use cuda:0
62%		729/1167	[01:33<00:57,	7.61it/s]	Device set to use cuda:0
63%		730/1167	[01:33<00:57,	7.65it/s]	Device set to use cuda:0
63%		731/1167	[01:33<00:55,	7.82it/s]	Device set to use cuda:0
63%		732/1167	[01:33<00:55,	7.83it/s]	Device set to use cuda:0
63%		733/1167	[01:33<00:54,	7.97it/s]	Device set to use cuda:0
63%		734/1167	[01:33<00:54,	7.90it/s]	Device set to use cuda:0
63%		735/1167	[01:33<00:54,	7.96it/s]	Device set to use cuda:0
63%		736/1167	[01:34<00:53,	8.03it/s]	Device set to use cuda:0
63%		737/1167	[01:34<00:53,	8.04it/s]	Device set to use cuda:0
63%		738/1167	[01:34<00:53,	8.06it/s]	Device set to use cuda:0
63%		739/1167	[01:34<00:52,	8.10it/s]	Device set to use cuda:0
63%		740/1167	[01:34<00:52,	8.07it/s]	Device set to use cuda:0
63%		741/1167	[01:34<00:52,	8.17it/s]	Device set to use cuda:0
64%		742/1167	[01:34<00:51,	8.20it/s]	Device set to use cuda:0
64%		743/1167	[01:34<00:57,	7.40it/s]	Device set to use cuda:0
64%		744/1167	[01:35<01:01,	6.88it/s]	Device set to use cuda:0
64%		745/1167	[01:35<00:58,	7.18it/s]	Device set to use cuda:0
64%		746/1167	[01:35<00:59,	7.12it/s]	Device set to use cuda:0
64%		747/1167	[01:35<00:59,	7.05it/s]	Device set to use cuda:0
64%		748/1167	[01:35<00:57,	7.25it/s]	Device set to use cuda:0
64%		749/1167	[01:35<00:56,	7.46it/s]	Device set to use cuda:0
64%		750/1167	[01:35<00:55,	7.57it/s]	Device set to use cuda:0
64%		751/1167	[01:36<00:53,	7.72it/s]	Device set to use cuda:0
64%		752/1167	[01:36<00:53,	7.83it/s]	Device set to use cuda:0
65%		753/1167	[01:36<00:53,	7.78it/s]	Device set to use cuda:0
65%		754/1167	[01:36<00:52,	7.79it/s]	Device set to use cuda:0
65%		755/1167	[01:36<00:51,	7.94it/s]	Device set to use cuda:0
65%		756/1167	[01:36<00:51,	8.06it/s]	Device set to use cuda:0
65%		757/1167	[01:36<00:51,	8.01it/s]	Device set to use cuda:0
65%		758/1167	[01:36<00:50,	8.09it/s]	Device set to use cuda:0
65%		759/1167	[01:37<00:50,	8.11it/s]	Device set to use cuda:0
65%		760/1167	[01:37<00:50,	8.05it/s]	Device set to use cuda:0
65%		761/1167	[01:37<00:50,	8.12it/s]	Device set to use cuda:0
65%		762/1167	[01:37<00:49,	8.20it/s]	Device set to use cuda:0
65%		763/1167	[01:37<00:49,	8.22it/s]	Device set to use cuda:0
65%		764/1167	[01:37<00:49,	8.19it/s]	Device set to use cuda:0
66%		765/1167	[01:37<00:52,	7.70it/s]	Device set to use cuda:0
66%		766/1167	[01:37<00:51,	7.74it/s]	Device set to use cuda:0
66%		767/1167	[01:38<00:51,	7.80it/s]	Device set to use cuda:0
66%		768/1167	[01:38<00:50,	7.94it/s]	Device set to use cuda:0
66%		769/1167	[01:38<00:50,	7.92it/s]	Device set to use cuda:0
66%		770/1167	[01:38<00:49,	7.98it/s]	Device set to use cuda:0
66%		771/1167	[01:38<00:48,	8.11it/s]	Device set to use cuda:0
66%		772/1167	[01:38<00:48,	8.15it/s]	Device set to use cuda:0
66%		773/1167	[01:38<00:48,	8.17it/s]	Device set to use cuda:0
66%		774/1167	[01:38<00:48,	8.14it/s]	Device set to use cuda:0
66%		775/1167	[01:39<00:48,	8.13it/s]	Device set to use cuda:0
66%		776/1167	[01:39<00:49,	7.92it/s]	Device set to use cuda:0
67%		777/1167	[01:39<00:48,	8.01it/s]	Device set to use cuda:0

67%		778/1167	[01:39<00:48,	8.04it/s]	Device set to use cuda:0
67%		779/1167	[01:39<00:47,	8.10it/s]	Device set to use cuda:0
67%		780/1167	[01:39<00:48,	7.99it/s]	Device set to use cuda:0
67%		781/1167	[01:39<00:48,	7.92it/s]	Device set to use cuda:0
67%		782/1167	[01:39<00:47,	8.04it/s]	Device set to use cuda:0
67%		783/1167	[01:40<00:49,	7.72it/s]	Device set to use cuda:0
67%		784/1167	[01:40<00:48,	7.95it/s]	Device set to use cuda:0
67%		785/1167	[01:40<00:48,	7.88it/s]	Device set to use cuda:0
67%		786/1167	[01:40<00:47,	8.00it/s]	Device set to use cuda:0
67%		787/1167	[01:40<00:47,	8.08it/s]	Device set to use cuda:0
68%		788/1167	[01:40<00:46,	8.07it/s]	Device set to use cuda:0
68%		789/1167	[01:40<00:46,	8.04it/s]	Device set to use cuda:0
68%		790/1167	[01:40<00:46,	8.16it/s]	Device set to use cuda:0
68%		791/1167	[01:41<00:46,	8.08it/s]	Device set to use cuda:0
68%		792/1167	[01:41<00:46,	8.11it/s]	Device set to use cuda:0
68%		793/1167	[01:41<00:45,	8.22it/s]	Device set to use cuda:0
68%		794/1167	[01:41<00:47,	7.91it/s]	Device set to use cuda:0
68%		795/1167	[01:41<00:46,	8.02it/s]	Device set to use cuda:0
68%		796/1167	[01:41<00:45,	8.13it/s]	Device set to use cuda:0
68%		797/1167	[01:41<00:45,	8.09it/s]	Device set to use cuda:0
68%		798/1167	[01:41<00:45,	8.11it/s]	Device set to use cuda:0
68%		799/1167	[01:42<00:46,	7.88it/s]	Device set to use cuda:0
69%		800/1167	[01:42<00:48,	7.51it/s]	Device set to use cuda:0
69%		801/1167	[01:42<00:47,	7.68it/s]	Device set to use cuda:0
69%		802/1167	[01:42<00:46,	7.92it/s]	Device set to use cuda:0
69%		803/1167	[01:42<00:45,	8.00it/s]	Device set to use cuda:0
69%		804/1167	[01:42<00:44,	8.16it/s]	Device set to use cuda:0
69%		805/1167	[01:42<00:44,	8.06it/s]	Device set to use cuda:0
69%		806/1167	[01:42<00:44,	8.12it/s]	Device set to use cuda:0
69%		807/1167	[01:43<00:44,	8.04it/s]	Device set to use cuda:0
69%		808/1167	[01:43<00:44,	8.06it/s]	Device set to use cuda:0
69%		809/1167	[01:43<00:44,	8.12it/s]	Device set to use cuda:0
69%		810/1167	[01:43<00:45,	7.90it/s]	Device set to use cuda:0
69%		811/1167	[01:43<00:44,	7.98it/s]	Device set to use cuda:0
70%		812/1167	[01:43<00:44,	8.03it/s]	Device set to use cuda:0
70%		813/1167	[01:43<00:43,	8.11it/s]	Device set to use cuda:0
70%		814/1167	[01:43<00:42,	8.22it/s]	Device set to use cuda:0
70%		815/1167	[01:44<00:43,	8.07it/s]	Device set to use cuda:0
70%		816/1167	[01:44<00:43,	8.05it/s]	Device set to use cuda:0
70%		817/1167	[01:44<00:46,	7.56it/s]	Device set to use cuda:0
70%		818/1167	[01:44<00:44,	7.76it/s]	Device set to use cuda:0
70%		819/1167	[01:44<00:44,	7.85it/s]	Device set to use cuda:0
70%		820/1167	[01:44<00:45,	7.66it/s]	Device set to use cuda:0
70%		821/1167	[01:44<00:44,	7.80it/s]	Device set to use cuda:0
70%		822/1167	[01:44<00:43,	7.95it/s]	Device set to use cuda:0
71%		823/1167	[01:45<00:42,	8.08it/s]	Device set to use cuda:0
71%		824/1167	[01:45<00:42,	8.14it/s]	Device set to use cuda:0
71%		825/1167	[01:45<00:41,	8.16it/s]	Device set to use cuda:0
71%		826/1167	[01:45<00:41,	8.19it/s]	Device set to use cuda:0
71%		827/1167	[01:45<00:41,	8.25it/s]	Device set to use cuda:0
71%		828/1167	[01:45<00:41,	8.23it/s]	Device set to use cuda:0
71%		829/1167	[01:45<00:41,	8.15it/s]	Device set to use cuda:0
71%		830/1167	[01:45<00:41,	8.21it/s]	Device set to use cuda:0
71%		831/1167	[01:46<00:41,	8.12it/s]	Device set to use cuda:0
71%		832/1167	[01:46<00:43,	7.70it/s]	Device set to use cuda:0
71%		833/1167	[01:46<00:42,	7.78it/s]	Device set to use cuda:0

71%			834/1167	[01:46<00:44,	7.46it/s]	Device set to use cuda:0
72%			835/1167	[01:46<00:43,	7.58it/s]	Device set to use cuda:0
72%			836/1167	[01:46<00:42,	7.70it/s]	Device set to use cuda:0
72%			837/1167	[01:46<00:42,	7.72it/s]	Device set to use cuda:0
72%			838/1167	[01:47<00:43,	7.57it/s]	Device set to use cuda:0
72%			839/1167	[01:47<00:42,	7.73it/s]	Device set to use cuda:0
72%			840/1167	[01:47<00:42,	7.69it/s]	Device set to use cuda:0
72%			841/1167	[01:47<00:41,	7.81it/s]	Device set to use cuda:0
72%			842/1167	[01:47<00:41,	7.90it/s]	Device set to use cuda:0
72%			843/1167	[01:47<00:40,	8.05it/s]	Device set to use cuda:0
72%			844/1167	[01:47<00:40,	8.05it/s]	Device set to use cuda:0
72%			845/1167	[01:47<00:42,	7.55it/s]	Device set to use cuda:0
72%			846/1167	[01:48<00:41,	7.74it/s]	Device set to use cuda:0
73%			847/1167	[01:48<00:41,	7.77it/s]	Device set to use cuda:0
73%			848/1167	[01:48<00:40,	7.93it/s]	Device set to use cuda:0
73%			849/1167	[01:48<00:40,	7.91it/s]	Device set to use cuda:0
73%			850/1167	[01:48<00:41,	7.72it/s]	Device set to use cuda:0
73%			851/1167	[01:48<00:41,	7.69it/s]	Device set to use cuda:0
73%			852/1167	[01:48<00:40,	7.87it/s]	Device set to use cuda:0
73%			853/1167	[01:48<00:39,	7.90it/s]	Device set to use cuda:0
73%			854/1167	[01:49<00:39,	7.97it/s]	Device set to use cuda:0
73%			855/1167	[01:49<00:39,	7.98it/s]	Device set to use cuda:0
73%			856/1167	[01:49<00:38,	8.05it/s]	Device set to use cuda:0
73%			857/1167	[01:49<00:38,	8.08it/s]	Device set to use cuda:0
74%			858/1167	[01:49<00:38,	8.07it/s]	Device set to use cuda:0
74%			859/1167	[01:49<00:37,	8.13it/s]	Device set to use cuda:0
74%			860/1167	[01:49<00:38,	8.04it/s]	Device set to use cuda:0
74%			861/1167	[01:49<00:39,	7.81it/s]	Device set to use cuda:0
74%			862/1167	[01:50<00:39,	7.74it/s]	Device set to use cuda:0
74%			863/1167	[01:50<00:39,	7.71it/s]	Device set to use cuda:0
74%			864/1167	[01:50<00:38,	7.84it/s]	Device set to use cuda:0
74%			865/1167	[01:50<00:38,	7.78it/s]	Device set to use cuda:0
74%			866/1167	[01:50<00:43,	6.89it/s]	Device set to use cuda:0
74%			867/1167	[01:50<00:43,	6.90it/s]	Device set to use cuda:0
74%			868/1167	[01:50<00:41,	7.13it/s]	Device set to use cuda:0
74%			869/1167	[01:51<00:40,	7.35it/s]	Device set to use cuda:0
75%			870/1167	[01:51<00:38,	7.64it/s]	Device set to use cuda:0
75%			871/1167	[01:51<00:38,	7.79it/s]	Device set to use cuda:0
75%			872/1167	[01:51<00:37,	7.87it/s]	Device set to use cuda:0
75%			873/1167	[01:51<00:36,	7.97it/s]	Device set to use cuda:0
75%			874/1167	[01:51<00:36,	8.01it/s]	Device set to use cuda:0
75%			875/1167	[01:51<00:36,	8.01it/s]	Device set to use cuda:0
75%			876/1167	[01:51<00:36,	7.95it/s]	Device set to use cuda:0
75%			877/1167	[01:51<00:36,	8.02it/s]	Device set to use cuda:0
75%			878/1167	[01:52<00:35,	8.14it/s]	Device set to use cuda:0
75%			879/1167	[01:52<00:35,	8.01it/s]	Device set to use cuda:0
75%			880/1167	[01:52<00:36,	7.97it/s]	Device set to use cuda:0
75%			881/1167	[01:52<00:35,	8.09it/s]	Device set to use cuda:0
76%			882/1167	[01:52<00:37,	7.65it/s]	Device set to use cuda:0
76%			883/1167	[01:52<00:36,	7.74it/s]	Device set to use cuda:0
76%			884/1167	[01:52<00:36,	7.84it/s]	Device set to use cuda:0
76%			885/1167	[01:53<00:35,	7.95it/s]	Device set to use cuda:0
76%			886/1167	[01:53<00:34,	8.10it/s]	Device set to use cuda:0
76%			887/1167	[01:53<00:34,	8.06it/s]	Device set to use cuda:0
76%			888/1167	[01:53<00:34,	8.12it/s]	Device set to use cuda:0
76%			889/1167	[01:53<00:33,	8.19it/s]	Device set to use cuda:0

76%		890/1167	[01:53<00:33,	8.21it/s]	Device set to use cuda:0
76%		891/1167	[01:53<00:33,	8.20it/s]	Device set to use cuda:0
76%		892/1167	[01:53<00:34,	8.09it/s]	Device set to use cuda:0
77%		893/1167	[01:54<00:35,	7.69it/s]	Device set to use cuda:0
77%		894/1167	[01:54<00:34,	7.84it/s]	Device set to use cuda:0
77%		895/1167	[01:54<00:34,	7.82it/s]	Device set to use cuda:0
77%		896/1167	[01:54<00:34,	7.85it/s]	Device set to use cuda:0
77%		897/1167	[01:54<00:33,	8.06it/s]	Device set to use cuda:0
77%		898/1167	[01:54<00:34,	7.69it/s]	Device set to use cuda:0
77%		899/1167	[01:54<00:34,	7.88it/s]	Device set to use cuda:0
77%		900/1167	[01:54<00:34,	7.78it/s]	Device set to use cuda:0
77%		901/1167	[01:55<00:34,	7.82it/s]	Device set to use cuda:0
77%		902/1167	[01:55<00:33,	7.97it/s]	Device set to use cuda:0
77%		903/1167	[01:55<00:32,	8.04it/s]	Device set to use cuda:0
77%		904/1167	[01:55<00:32,	8.06it/s]	Device set to use cuda:0
78%		905/1167	[01:55<00:32,	8.17it/s]	Device set to use cuda:0
78%		906/1167	[01:55<00:31,	8.21it/s]	Device set to use cuda:0
78%		907/1167	[01:55<00:31,	8.26it/s]	Device set to use cuda:0
78%		908/1167	[01:55<00:31,	8.12it/s]	Device set to use cuda:0
78%		909/1167	[01:55<00:31,	8.11it/s]	Device set to use cuda:0
78%		910/1167	[01:56<00:31,	8.10it/s]	Device set to use cuda:0
78%		911/1167	[01:56<00:32,	7.98it/s]	Device set to use cuda:0
78%		912/1167	[01:56<00:31,	8.00it/s]	Device set to use cuda:0
78%		913/1167	[01:56<00:31,	8.12it/s]	Device set to use cuda:0
78%		914/1167	[01:56<00:31,	8.15it/s]	Device set to use cuda:0
78%		915/1167	[01:56<00:30,	8.25it/s]	Device set to use cuda:0
78%		916/1167	[01:56<00:32,	7.66it/s]	Device set to use cuda:0
79%		917/1167	[01:57<00:31,	7.85it/s]	Device set to use cuda:0
79%		918/1167	[01:57<00:31,	7.92it/s]	Device set to use cuda:0
79%		919/1167	[01:57<00:31,	7.93it/s]	Device set to use cuda:0
79%		920/1167	[01:57<00:33,	7.45it/s]	Device set to use cuda:0
79%		921/1167	[01:57<00:32,	7.68it/s]	Device set to use cuda:0
79%		922/1167	[01:57<00:31,	7.87it/s]	Device set to use cuda:0
79%		923/1167	[01:57<00:30,	7.91it/s]	Device set to use cuda:0
79%		924/1167	[01:57<00:30,	8.00it/s]	Device set to use cuda:0
79%		925/1167	[01:58<00:29,	8.15it/s]	Device set to use cuda:0
79%		926/1167	[01:58<00:29,	8.21it/s]	Device set to use cuda:0
79%		927/1167	[01:58<00:29,	8.24it/s]	Device set to use cuda:0
80%		928/1167	[01:58<00:29,	8.15it/s]	Device set to use cuda:0
80%		929/1167	[01:58<00:29,	8.12it/s]	Device set to use cuda:0
80%		930/1167	[01:58<00:28,	8.22it/s]	Device set to use cuda:0
80%		931/1167	[01:58<00:28,	8.26it/s]	Device set to use cuda:0
80%		932/1167	[01:58<00:28,	8.12it/s]	Device set to use cuda:0
80%		933/1167	[01:58<00:28,	8.20it/s]	Device set to use cuda:0
80%		934/1167	[01:59<00:28,	8.17it/s]	Device set to use cuda:0
80%		935/1167	[01:59<00:29,	7.99it/s]	Device set to use cuda:0
80%		936/1167	[01:59<00:30,	7.51it/s]	Device set to use cuda:0
80%		937/1167	[01:59<00:31,	7.20it/s]	Device set to use cuda:0
80%		938/1167	[01:59<00:30,	7.48it/s]	Device set to use cuda:0
80%		939/1167	[01:59<00:29,	7.68it/s]	Device set to use cuda:0
81%		940/1167	[01:59<00:28,	7.91it/s]	Device set to use cuda:0
81%		941/1167	[02:00<00:28,	7.98it/s]	Device set to use cuda:0
81%		942/1167	[02:00<00:27,	8.10it/s]	Device set to use cuda:0
81%		943/1167	[02:00<00:27,	8.13it/s]	Device set to use cuda:0
81%		944/1167	[02:00<00:27,	8.00it/s]	Device set to use cuda:0
81%		945/1167	[02:00<00:27,	7.95it/s]	Device set to use cuda:0

81%			946/1167	[02:00<00:27,	7.97it/s]	Device set to use cuda:0
81%			947/1167	[02:00<00:27,	8.02it/s]	Device set to use cuda:0
81%			948/1167	[02:00<00:27,	8.03it/s]	Device set to use cuda:0
81%			949/1167	[02:01<00:27,	7.99it/s]	Device set to use cuda:0
81%			950/1167	[02:01<00:26,	8.12it/s]	Device set to use cuda:0
81%			951/1167	[02:01<00:26,	8.17it/s]	Device set to use cuda:0
82%			952/1167	[02:01<00:26,	8.10it/s]	Device set to use cuda:0
82%			953/1167	[02:01<00:26,	8.15it/s]	Device set to use cuda:0
82%			954/1167	[02:01<00:26,	8.13it/s]	Device set to use cuda:0
82%			955/1167	[02:01<00:27,	7.80it/s]	Device set to use cuda:0
82%			956/1167	[02:01<00:27,	7.77it/s]	Device set to use cuda:0
82%			957/1167	[02:02<00:26,	7.86it/s]	Device set to use cuda:0
82%			958/1167	[02:02<00:26,	7.97it/s]	Device set to use cuda:0
82%			959/1167	[02:02<00:26,	7.99it/s]	Device set to use cuda:0
82%			960/1167	[02:02<00:26,	7.70it/s]	Device set to use cuda:0
82%			961/1167	[02:02<00:26,	7.78it/s]	Device set to use cuda:0
82%			962/1167	[02:02<00:25,	7.93it/s]	Device set to use cuda:0
83%			963/1167	[02:02<00:25,	7.95it/s]	Device set to use cuda:0
83%			964/1167	[02:02<00:25,	7.99it/s]	Device set to use cuda:0
83%			965/1167	[02:03<00:25,	7.97it/s]	Device set to use cuda:0
83%			966/1167	[02:03<00:24,	8.06it/s]	Device set to use cuda:0
83%			967/1167	[02:03<00:24,	8.10it/s]	Device set to use cuda:0

Unsloth: Input IDs of length 1228 > the model's max sequence length of 1024.
We shall truncate it ourselves. It's imperative if you correct this issue first.

83%			968/1167	[02:03<00:27,	7.11it/s]	Device set to use cuda:0
83%			969/1167	[02:03<00:26,	7.38it/s]	Device set to use cuda:0
83%			970/1167	[02:03<00:25,	7.62it/s]	Device set to use cuda:0
83%			971/1167	[02:03<00:25,	7.74it/s]	Device set to use cuda:0
83%			972/1167	[02:03<00:24,	7.84it/s]	Device set to use cuda:0
83%			973/1167	[02:04<00:25,	7.52it/s]	Device set to use cuda:0
83%			974/1167	[02:04<00:24,	7.81it/s]	Device set to use cuda:0
84%			975/1167	[02:04<00:24,	7.88it/s]	Device set to use cuda:0
84%			976/1167	[02:04<00:24,	7.82it/s]	Device set to use cuda:0
84%			977/1167	[02:04<00:23,	7.94it/s]	Device set to use cuda:0
84%			978/1167	[02:04<00:23,	8.04it/s]	Device set to use cuda:0
84%			979/1167	[02:04<00:23,	8.04it/s]	Device set to use cuda:0
84%			980/1167	[02:04<00:23,	8.11it/s]	Device set to use cuda:0
84%			981/1167	[02:05<00:22,	8.12it/s]	Device set to use cuda:0
84%			982/1167	[02:05<00:22,	8.16it/s]	Device set to use cuda:0
84%			983/1167	[02:05<00:22,	8.21it/s]	Device set to use cuda:0
84%			984/1167	[02:05<00:22,	8.21it/s]	Device set to use cuda:0
84%			985/1167	[02:05<00:22,	8.23it/s]	Device set to use cuda:0
84%			986/1167	[02:05<00:21,	8.23it/s]	Device set to use cuda:0
85%			987/1167	[02:05<00:21,	8.19it/s]	Device set to use cuda:0
85%			988/1167	[02:05<00:21,	8.16it/s]	Device set to use cuda:0
85%			989/1167	[02:06<00:21,	8.21it/s]	Device set to use cuda:0
85%			990/1167	[02:06<00:21,	8.21it/s]	Device set to use cuda:0
85%			991/1167	[02:06<00:23,	7.40it/s]	Device set to use cuda:0
85%			992/1167	[02:06<00:23,	7.48it/s]	Device set to use cuda:0
85%			993/1167	[02:06<00:22,	7.66it/s]	Device set to use cuda:0
85%			994/1167	[02:06<00:21,	7.90it/s]	Device set to use cuda:0
85%			995/1167	[02:06<00:21,	8.11it/s]	Device set to use cuda:0
85%			996/1167	[02:06<00:21,	8.14it/s]	Device set to use cuda:0
85%			997/1167	[02:07<00:21,	8.05it/s]	Device set to use cuda:0
86%			998/1167	[02:07<00:20,	8.06it/s]	Device set to use cuda:0

86%		999/1167	[02:07<00:20,	8.09it/s]	Device set to use cuda:0
86%		1000/1167	[02:07<00:20,	8.15it/s]	Device set to use cuda:0
86%		1001/1167	[02:07<00:20,	8.13it/s]	Device set to use cuda:0
86%		1002/1167	[02:07<00:20,	8.12it/s]	Device set to use cuda:0
86%		1003/1167	[02:07<00:20,	8.20it/s]	Device set to use cuda:0
86%		1004/1167	[02:07<00:19,	8.20it/s]	Device set to use cuda:0
86%		1005/1167	[02:08<00:19,	8.29it/s]	Device set to use cuda:0
86%		1006/1167	[02:08<00:19,	8.30it/s]	Device set to use cuda:0
86%		1007/1167	[02:08<00:19,	8.33it/s]	Device set to use cuda:0
86%		1008/1167	[02:08<00:19,	8.37it/s]	Device set to use cuda:0
86%		1009/1167	[02:08<00:21,	7.48it/s]	Device set to use cuda:0
87%		1010/1167	[02:08<00:20,	7.49it/s]	Device set to use cuda:0
87%		1011/1167	[02:08<00:20,	7.70it/s]	Device set to use cuda:0
87%		1012/1167	[02:08<00:20,	7.74it/s]	Device set to use cuda:0
87%		1013/1167	[02:09<00:19,	7.75it/s]	Device set to use cuda:0
87%		1014/1167	[02:09<00:19,	7.85it/s]	Device set to use cuda:0
87%		1015/1167	[02:09<00:18,	8.02it/s]	Device set to use cuda:0
87%		1016/1167	[02:09<00:19,	7.80it/s]	Device set to use cuda:0
87%		1017/1167	[02:09<00:19,	7.88it/s]	Device set to use cuda:0
87%		1018/1167	[02:09<00:18,	7.99it/s]	Device set to use cuda:0
87%		1019/1167	[02:09<00:18,	8.00it/s]	Device set to use cuda:0
87%		1020/1167	[02:09<00:18,	8.06it/s]	Device set to use cuda:0
87%		1021/1167	[02:10<00:17,	8.14it/s]	Device set to use cuda:0
88%		1022/1167	[02:10<00:17,	8.13it/s]	Device set to use cuda:0
88%		1023/1167	[02:10<00:17,	8.24it/s]	Device set to use cuda:0
88%		1024/1167	[02:10<00:17,	8.25it/s]	Device set to use cuda:0
88%		1025/1167	[02:10<00:17,	7.94it/s]	Device set to use cuda:0
88%		1026/1167	[02:10<00:17,	7.96it/s]	Device set to use cuda:0
88%		1027/1167	[02:10<00:18,	7.65it/s]	Device set to use cuda:0
88%		1028/1167	[02:10<00:17,	7.80it/s]	Device set to use cuda:0
88%		1029/1167	[02:11<00:17,	7.80it/s]	Device set to use cuda:0
88%		1030/1167	[02:11<00:17,	7.95it/s]	Device set to use cuda:0
88%		1031/1167	[02:11<00:17,	7.99it/s]	Device set to use cuda:0
88%		1032/1167	[02:11<00:16,	8.07it/s]	Device set to use cuda:0
89%		1033/1167	[02:11<00:16,	8.13it/s]	Device set to use cuda:0
89%		1034/1167	[02:11<00:16,	8.22it/s]	Device set to use cuda:0
89%		1035/1167	[02:11<00:15,	8.28it/s]	Device set to use cuda:0
89%		1036/1167	[02:11<00:15,	8.30it/s]	Device set to use cuda:0
89%		1037/1167	[02:12<00:15,	8.39it/s]	Device set to use cuda:0
89%		1038/1167	[02:12<00:15,	8.42it/s]	Device set to use cuda:0
89%		1039/1167	[02:12<00:15,	8.39it/s]	Device set to use cuda:0
89%		1040/1167	[02:12<00:15,	8.45it/s]	Device set to use cuda:0
89%		1041/1167	[02:12<00:15,	8.31it/s]	Device set to use cuda:0
89%		1042/1167	[02:12<00:15,	8.19it/s]	Device set to use cuda:0
89%		1043/1167	[02:12<00:15,	8.20it/s]	Device set to use cuda:0
89%		1044/1167	[02:12<00:15,	8.15it/s]	Device set to use cuda:0
90%		1045/1167	[02:13<00:15,	7.77it/s]	Device set to use cuda:0
90%		1046/1167	[02:13<00:15,	7.90it/s]	Device set to use cuda:0
90%		1047/1167	[02:13<00:14,	8.00it/s]	Device set to use cuda:0
90%		1048/1167	[02:13<00:14,	8.03it/s]	Device set to use cuda:0
90%		1049/1167	[02:13<00:14,	8.11it/s]	Device set to use cuda:0
90%		1050/1167	[02:13<00:14,	8.11it/s]	Device set to use cuda:0
90%		1051/1167	[02:13<00:14,	8.11it/s]	Device set to use cuda:0
90%		1052/1167	[02:13<00:14,	7.81it/s]	Device set to use cuda:0
90%		1053/1167	[02:14<00:14,	7.86it/s]	Device set to use cuda:0
90%		1054/1167	[02:14<00:14,	7.83it/s]	Device set to use cuda:0

90%		1055/1167	[02:14<00:14,	7.85it/s]	Device set to use cuda:0
90%		1056/1167	[02:14<00:13,	8.02it/s]	Device set to use cuda:0
91%		1057/1167	[02:14<00:13,	7.99it/s]	Device set to use cuda:0
91%		1058/1167	[02:14<00:13,	7.85it/s]	Device set to use cuda:0
91%		1059/1167	[02:14<00:13,	7.87it/s]	Device set to use cuda:0
91%		1060/1167	[02:14<00:13,	7.79it/s]	Device set to use cuda:0
91%		1061/1167	[02:15<00:13,	7.87it/s]	Device set to use cuda:0
91%		1062/1167	[02:15<00:13,	7.93it/s]	Device set to use cuda:0
91%		1063/1167	[02:15<00:12,	8.04it/s]	Device set to use cuda:0
91%		1064/1167	[02:15<00:13,	7.60it/s]	Device set to use cuda:0
91%		1065/1167	[02:15<00:13,	7.71it/s]	Device set to use cuda:0
91%		1066/1167	[02:15<00:12,	7.82it/s]	Device set to use cuda:0
91%		1067/1167	[02:15<00:12,	7.91it/s]	Device set to use cuda:0
92%		1068/1167	[02:15<00:12,	7.98it/s]	Device set to use cuda:0
92%		1069/1167	[02:16<00:12,	7.88it/s]	Device set to use cuda:0
92%		1070/1167	[02:16<00:12,	7.94it/s]	Device set to use cuda:0
92%		1071/1167	[02:16<00:12,	7.97it/s]	Device set to use cuda:0
92%		1072/1167	[02:16<00:11,	7.99it/s]	Device set to use cuda:0
92%		1073/1167	[02:16<00:11,	7.96it/s]	Device set to use cuda:0
92%		1074/1167	[02:16<00:11,	7.95it/s]	Device set to use cuda:0
92%		1075/1167	[02:16<00:11,	8.07it/s]	Device set to use cuda:0
92%		1076/1167	[02:16<00:11,	8.00it/s]	Device set to use cuda:0
92%		1077/1167	[02:17<00:11,	8.12it/s]	Device set to use cuda:0
92%		1078/1167	[02:17<00:10,	8.12it/s]	Device set to use cuda:0
92%		1079/1167	[02:17<00:10,	8.23it/s]	Device set to use cuda:0
93%		1080/1167	[02:17<00:10,	8.21it/s]	Device set to use cuda:0
93%		1081/1167	[02:17<00:10,	8.24it/s]	Device set to use cuda:0
93%		1082/1167	[02:17<00:10,	7.80it/s]	Device set to use cuda:0
93%		1083/1167	[02:17<00:10,	7.92it/s]	Device set to use cuda:0
93%		1084/1167	[02:17<00:10,	8.12it/s]	Device set to use cuda:0
93%		1085/1167	[02:18<00:09,	8.21it/s]	Device set to use cuda:0
93%		1086/1167	[02:18<00:09,	8.17it/s]	Device set to use cuda:0
93%		1087/1167	[02:18<00:09,	8.11it/s]	Device set to use cuda:0
93%		1088/1167	[02:18<00:09,	8.17it/s]	Device set to use cuda:0
93%		1089/1167	[02:18<00:09,	8.12it/s]	Device set to use cuda:0
93%		1090/1167	[02:18<00:09,	8.07it/s]	Device set to use cuda:0
93%		1091/1167	[02:18<00:09,	8.05it/s]	Device set to use cuda:0
94%		1092/1167	[02:18<00:09,	7.97it/s]	Device set to use cuda:0
94%		1093/1167	[02:19<00:09,	8.06it/s]	Device set to use cuda:0
94%		1094/1167	[02:19<00:08,	8.12it/s]	Device set to use cuda:0
94%		1095/1167	[02:19<00:08,	8.14it/s]	Device set to use cuda:0
94%		1096/1167	[02:19<00:08,	8.11it/s]	Device set to use cuda:0
94%		1097/1167	[02:19<00:08,	8.14it/s]	Device set to use cuda:0
94%		1098/1167	[02:19<00:08,	8.15it/s]	Device set to use cuda:0
94%		1099/1167	[02:19<00:08,	8.25it/s]	Device set to use cuda:0
94%		1100/1167	[02:19<00:08,	7.71it/s]	Device set to use cuda:0
94%		1101/1167	[02:20<00:08,	7.86it/s]	Device set to use cuda:0
94%		1102/1167	[02:20<00:08,	8.04it/s]	Device set to use cuda:0
95%		1103/1167	[02:20<00:07,	8.14it/s]	Device set to use cuda:0
95%		1104/1167	[02:20<00:07,	8.12it/s]	Device set to use cuda:0
95%		1105/1167	[02:20<00:07,	7.96it/s]	Device set to use cuda:0
95%		1106/1167	[02:20<00:07,	7.95it/s]	Device set to use cuda:0
95%		1107/1167	[02:20<00:07,	7.98it/s]	Device set to use cuda:0
95%		1108/1167	[02:20<00:07,	7.77it/s]	Device set to use cuda:0
95%		1109/1167	[02:21<00:07,	7.87it/s]	Device set to use cuda:0
95%		1110/1167	[02:21<00:07,	7.81it/s]	Device set to use cuda:0

95%		1111/1167	[02:21<00:07,	7.85it/s]	Device set to use cuda:0
95%		1112/1167	[02:21<00:07,	7.68it/s]	Device set to use cuda:0
95%		1113/1167	[02:21<00:06,	7.74it/s]	Device set to use cuda:0
95%		1114/1167	[02:21<00:07,	7.53it/s]	Device set to use cuda:0
96%		1115/1167	[02:21<00:06,	7.57it/s]	Device set to use cuda:0
96%		1116/1167	[02:22<00:07,	7.11it/s]	Device set to use cuda:0
96%		1117/1167	[02:22<00:06,	7.24it/s]	Device set to use cuda:0
96%		1118/1167	[02:22<00:06,	7.34it/s]	Device set to use cuda:0
96%		1119/1167	[02:22<00:06,	7.51it/s]	Device set to use cuda:0
96%		1120/1167	[02:22<00:06,	7.51it/s]	Device set to use cuda:0
96%		1121/1167	[02:22<00:05,	7.69it/s]	Device set to use cuda:0
96%		1122/1167	[02:22<00:05,	7.66it/s]	Device set to use cuda:0
96%		1123/1167	[02:22<00:05,	7.85it/s]	Device set to use cuda:0
96%		1124/1167	[02:23<00:05,	7.86it/s]	Device set to use cuda:0
96%		1125/1167	[02:23<00:05,	7.94it/s]	Device set to use cuda:0
96%		1126/1167	[02:23<00:05,	7.58it/s]	Device set to use cuda:0
97%		1127/1167	[02:23<00:05,	7.80it/s]	Device set to use cuda:0
97%		1128/1167	[02:23<00:04,	7.91it/s]	Device set to use cuda:0
97%		1129/1167	[02:23<00:04,	8.01it/s]	Device set to use cuda:0
97%		1130/1167	[02:23<00:04,	8.15it/s]	Device set to use cuda:0
97%		1131/1167	[02:23<00:04,	7.85it/s]	Device set to use cuda:0
97%		1132/1167	[02:24<00:04,	7.96it/s]	Device set to use cuda:0
97%		1133/1167	[02:24<00:04,	8.07it/s]	Device set to use cuda:0
97%		1134/1167	[02:24<00:04,	8.13it/s]	Device set to use cuda:0
97%		1135/1167	[02:24<00:03,	8.18it/s]	Device set to use cuda:0
97%		1136/1167	[02:24<00:03,	8.24it/s]	Device set to use cuda:0
97%		1137/1167	[02:24<00:03,	8.12it/s]	Device set to use cuda:0
98%		1138/1167	[02:24<00:03,	8.15it/s]	Device set to use cuda:0
98%		1139/1167	[02:24<00:03,	8.11it/s]	Device set to use cuda:0
98%		1140/1167	[02:25<00:03,	8.06it/s]	Device set to use cuda:0
98%		1141/1167	[02:25<00:03,	7.88it/s]	Device set to use cuda:0
98%		1142/1167	[02:25<00:03,	7.77it/s]	Device set to use cuda:0
98%		1143/1167	[02:25<00:03,	7.67it/s]	Device set to use cuda:0
98%		1144/1167	[02:25<00:03,	7.25it/s]	Device set to use cuda:0
98%		1145/1167	[02:25<00:03,	6.95it/s]	Device set to use cuda:0
98%		1146/1167	[02:25<00:03,	6.83it/s]	Device set to use cuda:0
98%		1147/1167	[02:26<00:03,	6.40it/s]	Device set to use cuda:0
98%		1148/1167	[02:26<00:02,	6.53it/s]	Device set to use cuda:0
98%		1149/1167	[02:26<00:02,	6.37it/s]	Device set to use cuda:0
99%		1150/1167	[02:26<00:02,	6.26it/s]	Device set to use cuda:0
99%		1151/1167	[02:26<00:02,	5.79it/s]	Device set to use cuda:0
99%		1152/1167	[02:26<00:02,	5.88it/s]	Device set to use cuda:0
99%		1153/1167	[02:27<00:02,	5.68it/s]	Device set to use cuda:0
99%		1154/1167	[02:27<00:02,	5.80it/s]	Device set to use cuda:0
99%		1155/1167	[02:27<00:02,	5.70it/s]	Device set to use cuda:0
99%		1156/1167	[02:27<00:01,	5.99it/s]	Device set to use cuda:0
99%		1157/1167	[02:27<00:01,	6.17it/s]	Device set to use cuda:0
99%		1158/1167	[02:27<00:01,	5.82it/s]	Device set to use cuda:0
99%		1159/1167	[02:28<00:01,	6.01it/s]	Device set to use cuda:0
99%		1160/1167	[02:28<00:01,	5.98it/s]	Device set to use cuda:0
99%		1161/1167	[02:28<00:01,	5.96it/s]	Device set to use cuda:0
100%		1162/1167	[02:28<00:00,	6.18it/s]	Device set to use cuda:0
100%		1163/1167	[02:28<00:00,	6.35it/s]	Device set to use cuda:0
100%		1164/1167	[02:28<00:00,	6.37it/s]	Device set to use cuda:0
100%		1165/1167	[02:29<00:00,	6.63it/s]	Device set to use cuda:0

```
100%|██████████| 1166/1167 [02:29<00:00, 6.33it/s]Device set to use cuda:0
100%|██████████| 1167/1167 [02:29<00:00, 7.81it/s]
```

5. Evaluation of Pre-fine tuned Model

```
In [9]: def evaluate(y_true, y_pred):
        labels = ["Normal", "Depression", "Anxiety"]
        mapping = {label: idx for idx, label in enumerate(labels)}

        def map_func(x):
            return mapping.get(x, -1) # Map to -1 if not found, but should not

        y_true_mapped = np.vectorize(map_func)(y_true)
        y_pred_mapped = np.vectorize(map_func)(y_pred)

        # Calculate accuracy
        accuracy = accuracy_score(y_true=y_true_mapped, y_pred=y_pred_mapped)
        print(f'Accuracy: {accuracy:.3f}')

        # Generate accuracy report
        unique_labels = set(y_true_mapped) # Get unique labels

        for label in unique_labels:
            label_indices = [i for i in range(len(y_true_mapped)) if y_true_mapped[i] == label]
            label_y_true = [y_true_mapped[i] for i in label_indices]
            label_y_pred = [y_pred_mapped[i] for i in label_indices]
            label_accuracy = accuracy_score(label_y_true, label_y_pred)
            print(f'Accuracy for label {labels[label]}: {label_accuracy:.3f}')

        # Generate classification report
        class_report = classification_report(y_true=y_true_mapped, y_pred=y_pred_mapped)
        print('\nClassification Report:')
        print(class_report)

        # Generate confusion matrix
        conf_matrix = confusion_matrix(y_true=y_true_mapped, y_pred=y_pred_mapped)
        print('\nConfusion Matrix:')
        print(conf_matrix)

        evaluate(y_true, y_pred)
```

Accuracy: 0.730
Accuracy for label Normal: 0.746
Accuracy for label Depression: 0.853
Accuracy for label Anxiety: 0.591

Classification Report:

	precision	recall	f1-score	support
Normal	0.76	0.75	0.75	389
Depression	0.67	0.85	0.75	389
Anxiety	0.81	0.59	0.68	389
micro avg	0.73	0.73	0.73	1167
macro avg	0.75	0.73	0.73	1167
weighted avg	0.75	0.73	0.73	1167

Confusion Matrix:

```
[[290  82  17]
 [ 14 332  38]
 [ 76  83 230]]
```

6. Setting training and config parameters

```
In [10]: import bitsandbytes as bnb

def find_all_linear_names(model):
    cls = bnb.nn.Linear4bit
    lora_module_names = set()
    for name, module in model.named_modules():
        if isinstance(module, cls):
            names = name.split('.')
            lora_module_names.add(names[0] if len(names) == 1 else names[-1])
    if 'lm_head' in lora_module_names: # needed for 16 bit
        lora_module_names.remove('lm_head')
    return list(lora_module_names)
modules = find_all_linear_names(model)
modules
```

```
Out[10]: ['k_proj', 'q_proj', 'down_proj', 'gate_proj', 'o_proj', 'up_proj', 'v_proj']
```

```
In [11]: model = FastLanguageModel.get_peft_model(
    model,
    r = 16, # Choose any number > 0 ! Suggested 8, 16, 32, 64, 128
    target_modules = modules,
    lora_alpha = 16,
    lora_dropout = 0, # Supports any, but = 0 is optimized
    bias = "none", # Supports any, but = "none" is optimized
    # [NEW] "unsloth" uses 30% less VRAM, fits 2x larger batch sizes!
    use_gradient_checkpointing = "unsloth", # True or "unsloth" for very long
    random_state = 3407,
    use_rslora = False, # We support rank stabilized LoRA
)
```

Unsloth 2025.3.19 patched 32 layers with 32 QKV layers, 32 O layers and 32 MLP layers.

```
In [12]: # # 3. Optimize LoRA configuration - reduce rank and alpha
# peft_config = LoraConfig(
#     lora_alpha=8, # Reduced from 16
#     lora_dropout=0,
#     r=16, # Reduced from 64
#     bias="none",
#     task_type="CAUSAL_LM",
#     target_modules=modules,
# )

In [13]: output_dir="llama-3.1-fine-tuned-model"

# 4. Optimize training parameters
training_arguments = SFTConfig(
    output_dir=output_dir,
    num_train_epochs=1,
    per_device_train_batch_size=1,
    gradient_accumulation_steps=16, # Increased from 8
    gradient_checkpointing=True,
    optim="paged_adamw_8bit", # Use 8-bit optimizer to save memory
    logging_steps=20, # Reduced logging frequency
    learning_rate=2e-4,
    weight_decay=0.001,
    fp16=False,
    bf16=True,
    max_grad_norm=0.3,
    max_steps=-1,
    warmup_ratio=0.03,
    group_by_length=False,
    lr_scheduler_type="cosine",
    report_to="wandb",
    eval_strategy="steps",
    eval_steps=0.5, # Less frequent evaluation
    dataset_text_field="text",
    max_seq_length=256, # Reduced sequence length
    packing=False,
    dataset_kwargs={
        "add_special_tokens": False,
        "append_concat_token": False,
    },
)

# 5. Clear cache before training
import gc
import torch
gc.collect()
torch.cuda.empty_cache()

# Set up trainer
trainer = SFTTrainer(
    model=model,
    train_dataset=train_data,
```

```

eval_dataset=eval_data,
# peft_config=peft_config,
processing_class=tokenizer,
args=training_arguments,
)

```

```

Unsloth: Tokenizing ["text"] (num_proc=8): 0%|          | 0/9331 [00:00<?,
? examples/s]

```

```

Unsloth: Tokenizing ["text"] (num_proc=8): 0%|          | 0/1166 [00:00<?,
? examples/s]

```

7. Train the model

In [14]: *# Add manual garbage collection during training*

```

old_step = trainer.training_step

def training_step_with_gc(*args, **kwargs):
    if trainer.state.global_step % 50 == 0:
        gc.collect()
        torch.cuda.empty_cache()
    return old_step(*args, **kwargs)

trainer.training_step = training_step_with_gc

# Train and save
trainer_stats = trainer.train()
wandb.finish()
model.config.use_cache = True
trainer.save_model(output_dir)
tokenizer.save_pretrained(output_dir)

```

```

==(=====)== Unsloth - 2x faster free finetuning | Num GPUs used = 1
  \ \      /|   Num examples = 9,331 | Num Epochs = 1 | Total steps = 583
0^0/ \_/ \     Batch size per device = 1 | Gradient accumulation steps = 16
 \      /      Data Parallel GPUs = 1 | Total batch size (1 x 16 x 1) = 16
  "-_____"      Trainable parameters = 41,943,040/8,000,000,000 (0.52% traine
d)

```

wandb: **WARNING** The `run_name` is currently set to the same value as `TrainingArguments.output_dir`. If this was not intended, please specify a different run name by setting the `TrainingArguments.run_name` parameter.

[583/583 57:20, Epoch 0/1]

Step	Training Loss	Validation Loss
------	---------------	-----------------

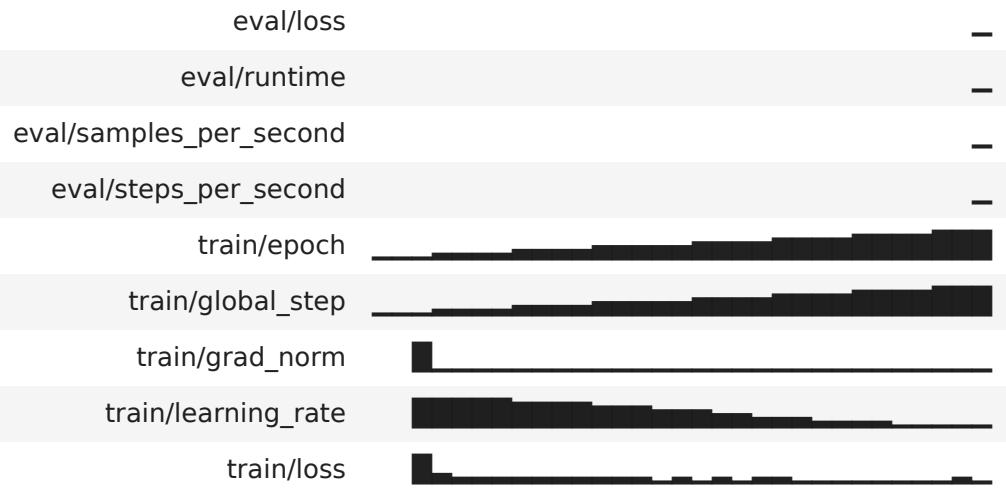
292	2.117200	2.042768
-----	----------	----------

Unsloth: Not an error, but LlamaForCausalLM does not accept `num_items_in_batch`.

Using gradient accumulation will be very slightly less accurate.

Read more on gradient accumulation issues here: <https://unsloth.ai/blog/gradient>

Run history:



Run summary:

eval/loss	2.04277
eval/runtime	42.8131
eval/samples_per_second	27.235
eval/steps_per_second	6.82
total_flos	5.316330584776704e+16
train/epoch	0.99968
train/global_step	583
train/grad_norm	0.40559
train/learning_rate	0.0
train/loss	2.0802
train_loss	2.13006
train_runtime	3452.8734
train_samples_per_second	2.702
train_steps_per_second	0.169

View run **dutiful-wind-19** at: <https://wandb.ai/dionyichia-nanyang-technological-university-singapore/Fine-tune%20llama-3.1-8b-it%20on%20Sentiment%20Analysis%20Dataset/runs/0oozrnc9?apiKey=73ba7c69e019dd2879de1bd311b39552ea00e93a>
View project at: <https://wandb.ai/dionyichia-nanyang-technological-university-singapore/Fine-tune%20llama-3.1-8b-it%20on%20Sentiment%20Analysis%20Dataset?apiKey=73ba7c69e019dd2879de1bd311b39552ea00e93a>
Synced 5 W&B file(s), 0 media file(s), 0 artifact file(s) and 0 other file(s)
Find logs at: ./wandb/run-20250417_163054-0oozrnc9/logs

```
Out[14]: ('llama-3.1-fine-tuned-model/tokenizer_config.json',  
         'llama-3.1-fine-tuned-model/special_tokens_map.json',  
         'llama-3.1-fine-tuned-model/tokenizer.json')
```

8. Evaluate trained model

```
In [15]: y_pred = predict(test_df, model, tokenizer)  
         evaluate(y_true, y_pred)
```

```
0%|          | 0/1167 [00:00<?, ?it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].
```

```
0%|          | 1/1167 [00:00<04:18, 4.52it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus
```

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

0%| | 2/1167 [00:00<03:40, 5.28it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

0%| | 3/1167 [00:00<03:28, 5.59it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

0%| | 4/1167 [00:00<03:27, 5.60it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

0%| | 5/1167 [00:00<03:32, 5.47it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

1%| | 6/1167 [00:01<03:25, 5.65it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

1%| | 7/1167 [00:01<03:23, 5.71it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

1%| | 8/1167 [00:01<03:18, 5.84it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

1%| | 9/1167 [00:01<03:31, 5.47it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

1%| | 10/1167 [00:01<03:26, 5.60it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

1%| | 11/1167 [00:01<03:22, 5.71it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

1%| | 12/1167 [00:02<03:22, 5.69it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

1%| | 13/1167 [00:02<03:20, 5.75it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

1%| | 14/1167 [00:02<03:17, 5.83it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

1%|| | 15/1167 [00:02<03:17, 5.82it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

1%|| | 16/1167 [00:02<03:16, 5.85it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

1%|| | 17/1167 [00:02<03:16, 5.86it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

2%|| | 18/1167 [00:03<03:13, 5.95it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

2%|| | 19/1167 [00:03<03:09, 6.04it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

2%|| | 20/1167 [00:03<03:08, 6.09it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

2%|| | 21/1167 [00:03<03:12, 5.97it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

2%|| | 22/1167 [00:03<03:15, 5.86it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

2%|| | 23/1167 [00:03<03:14, 5.88it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

2%|| | 24/1167 [00:04<03:17, 5.80it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

2%|| | 25/1167 [00:04<03:15, 5.83it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

2%|| | 26/1167 [00:04<03:15, 5.83it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

2%|| | 27/1167 [00:04<03:13, 5.88it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

2%|| | 28/1167 [00:04<03:31, 5.39it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

2%|| | 29/1167 [00:05<03:33, 5.32it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

3%|| | 30/1167 [00:05<03:26, 5.51it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

3%|| | 31/1167 [00:05<03:21, 5.62it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

3%|| | 32/1167 [00:05<03:19, 5.69it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

3%|| | 33/1167 [00:05<03:20, 5.65it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

3%|| | 34/1167 [00:05<03:18, 5.72it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

3%|| | 35/1167 [00:06<03:15, 5.79it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

3%|| | 36/1167 [00:06<03:16, 5.75it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

3%|| | 37/1167 [00:06<03:18, 5.69it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

3%|| | 38/1167 [00:06<03:17, 5.71it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

3%|| | 39/1167 [00:06<03:15, 5.77it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

3%|| | 40/1167 [00:06<03:11, 5.88it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

4%|| | 41/1167 [00:07<03:11, 5.87it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

4%|| | 42/1167 [00:07<03:10, 5.91it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

4%|| | 43/1167 [00:07<03:26, 5.43it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

4%| | 44/1167 [00:07<03:31, 5.31it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

4%|| | 45/1167 [00:07<03:26, 5.45it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

4%|| | 46/1167 [00:08<03:18, 5.64it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

4%|| | 47/1167 [00:08<03:19, 5.61it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

4%|| | 48/1167 [00:08<03:15, 5.72it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

4%|| | 49/1167 [00:08<03:12, 5.80it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

4%| | 50/1167 [00:08<03:13, 5.77it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

4%| | 51/1167 [00:08<03:09, 5.88it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

4%|| | 52/1167 [00:09<03:07, 5.94it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

5%| | 53/1167 [00:09<03:07, 5.95it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

5%| | 54/1167 [00:09<03:13, 5.75it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

5%|| | 55/1167 [00:09<03:10, 5.84it/s]Device set to use cuda:0
The model 'PefitModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

5%|| | 56/1167 [00:09<03:10, 5.84it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

5%|| | 57/1167 [00:09<03:08, 5.88it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

5%| | 58/1167 [00:10<03:09, 5.84it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

5%| | 59/1167 [00:10<03:11, 5.79it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

5%| | 60/1167 [00:10<03:08, 5.88it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

5%| | 61/1167 [00:10<03:05, 5.97it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

5%| | 62/1167 [00:10<03:06, 5.93it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

5%| | 63/1167 [00:10<03:05, 5.94it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

5%| | 64/1167 [00:11<03:12, 5.74it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

6%| | 65/1167 [00:11<03:07, 5.88it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

6%| | 66/1167 [00:11<03:03, 5.99it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

6%| | 67/1167 [00:11<03:02, 6.04it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

6%| | 68/1167 [00:11<03:10, 5.76it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

6%| | 69/1167 [00:11<03:07, 5.87it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

6%| | 70/1167 [00:12<03:04, 5.94it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

6%| | 71/1167 [00:12<03:04, 5.94it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

6%| | 72/1167 [00:12<03:00, 6.07it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

6%| | 73/1167 [00:12<03:01, 6.03it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

6%| | 74/1167 [00:12<03:10, 5.73it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

6%| | 75/1167 [00:13<03:07, 5.83it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

7%|█ | 76/1167 [00:13<03:06, 5.84it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

7%| | 77/1167 [00:13<03:03, 5.94it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

7%| | 78/1167 [00:13<03:00, 6.02it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

7%|█ | 79/1167 [00:13<02:59, 6.06it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

7%| | 80/1167 [00:13<03:03, 5.93it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

7%| | 81/1167 [00:13<03:00, 6.01it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

7%|█ | 82/1167 [00:14<03:02, 5.94it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

7%|█ | 83/1167 [00:14<03:15, 5.56it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

7%|█ | 84/1167 [00:14<03:07, 5.76it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

7%|█| 85/1167 [00:14<03:03, 5.90it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

7%|█| 86/1167 [00:14<03:03, 5.88it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

7%| | 87/1167 [00:15<03:01, 5.94it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

8%| | 88/1167 [00:15<03:02, 5.90it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

8%| | 89/1167 [00:15<03:03, 5.88it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

8%|█ | 90/1167 [00:15<03:01, 5.94it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

8%|█ | 91/1167 [00:15<03:02, 5.89it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

8%|██████████| 92/1167 [00:15<03:12, 5.59it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

8%|█ | 93/1167 [00:16<03:08, 5.71it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

8%|█ | 94/1167 [00:16<03:06, 5.75it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

8%|█ | 95/1167 [00:16<03:07, 5.72it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

8%| | 96/1167 [00:16<03:02, 5.86it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

8%| | 97/1167 [00:16<03:03, 5.84it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

8%|█ | 98/1167 [00:16<03:02, 5.86it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

8%|█ | 99/1167 [00:17<03:00, 5.91it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

9%|█ | 100/1167 [00:17<03:09, 5.62it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

9%|█ | 101/1167 [00:17<03:19, 5.33it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

9%|█ | 102/1167 [00:17<03:12, 5.53it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

9%|█ | 103/1167 [00:17<03:13, 5.51it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

9%|█ | 104/1167 [00:18<03:06, 5.71it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

9%|█ | 105/1167 [00:18<03:10, 5.57it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

9%|█ | 106/1167 [00:18<03:11, 5.54it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

9%|█ | 107/1167 [00:18<03:04, 5.74it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

9%|█ | 108/1167 [00:18<03:02, 5.79it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

9%|██████████| 109/1167 [00:18<03:02, 5.79it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

9%|██████████| 110/1167 [00:19<03:01, 5.82it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

10%|█ | 111/1167 [00:19<03:14, 5.43it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

10%|█ | 112/1167 [00:19<03:15, 5.39it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

10%|█ | 113/1167 [00:19<03:13, 5.45it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

10%|█ | 114/1167 [00:19<03:15, 5.37it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

10%|█ | 115/1167 [00:20<03:11, 5.51it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

10%|█ | 116/1167 [00:20<03:19, 5.28it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

10%|██████████| 117/1167 [00:20<03:14, 5.39it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

10%|██████████| 118/1167 [00:20<03:20, 5.24it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

10%|██████████| 119/1167 [00:20<03:13, 5.41it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

10%|█ | 120/1167 [00:20<03:11, 5.48it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

10%|█ | 121/1167 [00:21<03:05, 5.64it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

10%|█ | 122/1167 [00:21<03:20, 5.22it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

11%|█ | 123/1167 [00:21<03:19, 5.23it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

11%|█ | 124/1167 [00:21<03:10, 5.47it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

11%|█| 125/1167 [00:21<03:12, 5.42it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

11%|█| 126/1167 [00:22<03:09, 5.50it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

11%|█ | 127/1167 [00:22<03:09, 5.49it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

11%|█ | 128/1167 [00:22<03:06, 5.56it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

11%|█ | 129/1167 [00:22<03:10, 5.46it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

11%|█ | 130/1167 [00:22<03:15, 5.30it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

11%|█ | 131/1167 [00:22<03:09, 5.47it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

11%|██████████| 132/1167 [00:23<03:10, 5.44it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

11%|█| 133/1167 [00:23<03:04, 5.60it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

11%|█| 134/1167 [00:23<03:02, 5.67it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

12%|██████████| 135/1167 [00:23<03:00, 5.72it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

12%|█ | 136/1167 [00:23<03:11, 5.39it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

12%|█ | 137/1167 [00:24<03:06, 5.52it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

12%|██████████| 138/1167 [00:24<03:04, 5.59it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

12%|██████████| 139/1167 [00:24<03:03, 5.59it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

12%|██████████| 140/1167 [00:24<03:01, 5.66it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

12%|█| 141/1167 [00:24<03:07, 5.47it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

12%|█| 142/1167 [00:24<03:04, 5.57it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

12%|██████████| 143/1167 [00:25<03:00, 5.68it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

12%|█ | 144/1167 [00:25<03:01, 5.62it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

12%|█ | 145/1167 [00:25<02:58, 5.73it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

13%|██████████| 146/1167 [00:25<03:01, 5.62it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

13%|██████████| 147/1167 [00:25<02:58, 5.70it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

13%|██████████| 148/1167 [00:25<02:56, 5.78it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

13%|██████████| 149/1167 [00:26<02:52, 5.90it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

13%|██████████| 150/1167 [00:26<03:05, 5.49it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

13%|██████████| 151/1167 [00:26<02:59, 5.67it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

13%|██████████ | 152/1167 [00:26<02:55, 5.78it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

13%|██████████ | 153/1167 [00:26<02:55, 5.78it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

13%|██████████| 154/1167 [00:27<02:52, 5.87it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

13%|██████████| 155/1167 [00:27<02:57, 5.70it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

13%|██████████| 156/1167 [00:27<02:52, 5.86it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

13%|██████████| 157/1167 [00:27<02:52, 5.87it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

14%|██████████| 158/1167 [00:27<02:51, 5.90it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

14%|██████████| 159/1167 [00:27<02:54, 5.78it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

14%|██████████ | 160/1167 [00:28<03:14, 5.19it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

14%|██████████ | 161/1167 [00:28<03:09, 5.30it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

14%|██████████| 162/1167 [00:28<03:07, 5.35it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

14%|██████████| 163/1167 [00:28<03:05, 5.41it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

14%|██████████| 164/1167 [00:28<03:05, 5.42it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

14%|██████████| 165/1167 [00:29<03:07, 5.33it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

14%|██████████| 166/1167 [00:29<03:04, 5.42it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

14%|██████████| 167/1167 [00:29<03:03, 5.44it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

14%|██████████ | 168/1167 [00:29<03:01, 5.50it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformersLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

14%|██████████ | 169/1167 [00:29<03:08, 5.29it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

15%|██████████| 170/1167 [00:29<03:04, 5.41it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

15%|██████████| 171/1167 [00:30<03:02, 5.45it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

15%|██████████| 172/1167 [00:30<03:02, 5.47it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

15%|██████████| 173/1167 [00:30<03:09, 5.24it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

15%|██████████| 174/1167 [00:30<03:02, 5.45it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

15%|██████████| 175/1167 [00:30<02:58, 5.56it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

15%|██████████ | 176/1167 [00:31<02:58, 5.56it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

15%|██████████ | 177/1167 [00:31<02:55, 5.63it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

15%|██████████| 178/1167 [00:31<02:52, 5.73it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

15%|██████████| 179/1167 [00:31<02:51, 5.76it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformersLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

15%|██████████| 180/1167 [00:31<02:52, 5.71it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

16%|██████████| 181/1167 [00:31<03:00, 5.48it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

16%|██████████| 182/1167 [00:32<02:58, 5.53it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

16%|██████████| 183/1167 [00:32<02:55, 5.59it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

16%|██████████ | 184/1167 [00:32<02:55, 5.61it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

16%|██████████ | 185/1167 [00:32<02:52, 5.70it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

16%|██████████| 186/1167 [00:32<02:53, 5.64it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

16%|██████████| 187/1167 [00:33<03:00, 5.43it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

16%|██████████| 188/1167 [00:33<02:58, 5.48it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

16%|██████████| 189/1167 [00:33<02:53, 5.64it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

16%|██████████| 190/1167 [00:33<02:50, 5.73it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

16%|██████████| 191/1167 [00:33<02:49, 5.77it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

16%|██████████ | 192/1167 [00:33<02:49, 5.77it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

17%|██████████ | 193/1167 [00:34<02:48, 5.78it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

17%|██████████| 194/1167 [00:34<02:50, 5.71it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

17%|██████████| 195/1167 [00:34<02:49, 5.72it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

17%|██████████| 196/1167 [00:34<02:54, 5.56it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

17%|██████████| 197/1167 [00:34<02:55, 5.53it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

17%|██████████| 198/1167 [00:34<02:54, 5.57it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

17%|██████████| 199/1167 [00:35<02:52, 5.6lit/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

17%|██████████| 200/1167 [00:35<02:58, 5.41it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

17%|██████████| 201/1167 [00:35<02:53, 5.57it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

17%|██████████| 202/1167 [00:35<02:51, 5.63it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

17%|██████████| 203/1167 [00:35<02:59, 5.36it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

17%|██████████| 204/1167 [00:36<02:55, 5.49it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

18%|██████████| 205/1167 [00:36<02:52, 5.57it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

18%|██████████| 206/1167 [00:36<02:52, 5.57it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

18%|██████████| 207/1167 [00:36<02:58, 5.37it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

18%|██████████ | 208/1167 [00:36<03:03, 5.24it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

18%|██████████ | 209/1167 [00:36<03:00, 5.32it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

18%|██████████| 210/1167 [00:37<02:54, 5.47it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

18%|██████████| 211/1167 [00:37<02:52, 5.55it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

18%|██████████| 212/1167 [00:37<02:51, 5.58it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

18%|██████████| 213/1167 [00:37<02:56, 5.39it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

18%|██████████| 214/1167 [00:37<02:53, 5.49it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

18%|██████████| 215/1167 [00:38<02:54, 5.45it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

19%|██████████ | 216/1167 [00:38<02:49, 5.60it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

19%|██████████ | 217/1167 [00:38<02:53, 5.48it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

19%|██████████| 218/1167 [00:38<02:51, 5.52it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

19%|██████████| 219/1167 [00:38<02:47, 5.66it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

19%|██████████| 220/1167 [00:38<02:47, 5.65it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

19%|██████████| 221/1167 [00:39<02:47, 5.64it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

19%|██████████| 222/1167 [00:39<02:51, 5.51it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

19%|██████████| 223/1167 [00:39<02:49, 5.58it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

19%|██████████| 224/1167 [00:39<02:51, 5.49it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

19%|██████████| 225/1167 [00:39<02:57, 5.31it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

19%|██████████| 226/1167 [00:40<02:53, 5.41it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

19%|██████████| 227/1167 [00:40<02:57, 5.29it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

20%|██████████| 228/1167 [00:40<02:52, 5.45it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

20%|██████████| 229/1167 [00:40<02:48, 5.55it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

20%|██████████| 230/1167 [00:40<02:48, 5.56it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

20%|██████████| 231/1167 [00:41<03:03, 5.11it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

20%|██████████ | 232/1167 [00:41<02:57, 5.28it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

20%|██████████ | 233/1167 [00:41<02:50, 5.48it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

20%|██████████| 234/1167 [00:41<02:45, 5.62it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

20%|██████████| 235/1167 [00:41<02:53, 5.38it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

20%|██████████| 236/1167 [00:41<02:50, 5.45it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

20%|██████████| 237/1167 [00:42<02:47, 5.57it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

20%|██████████| 238/1167 [00:42<02:45, 5.62it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

20%|██████████| 239/1167 [00:42<02:41, 5.73it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

21%|██████████ | 240/1167 [00:42<02:39, 5.81it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformersLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

21%|██████████ | 241/1167 [00:42<02:40, 5.75it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

21%|██████████| 242/1167 [00:42<02:42, 5.69it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

21%|██████████| 243/1167 [00:43<02:47, 5.53it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

21%|██████████| 244/1167 [00:43<02:44, 5.61it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

21%|██████████| 245/1167 [00:43<02:41, 5.71it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

21%|██████████| 246/1167 [00:43<02:38, 5.83it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

21%|██████████| 247/1167 [00:43<02:40, 5.73it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

21%|██████████ | 248/1167 [00:43<02:44, 5.59it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformersLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

21%|██████████ | 249/1167 [00:44<02:41, 5.70it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

21%|██████████| 250/1167 [00:44<02:37, 5.81it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

22%|██████████| 251/1167 [00:44<02:48, 5.43it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

22%|██████████| 252/1167 [00:44<02:45, 5.55it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

22%|██████████| 253/1167 [00:44<02:45, 5.53it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

22%|██████████| 254/1167 [00:45<02:44, 5.56it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

22%|██████████| 255/1167 [00:45<02:53, 5.26it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

22%|██████████ | 256/1167 [00:45<02:47, 5.45it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

22%|██████████ | 257/1167 [00:45<02:42, 5.59it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

22%|██████████| 258/1167 [00:45<02:39, 5.68it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

22%|██████████| 259/1167 [00:45<02:42, 5.59it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

22%|██████████| 260/1167 [00:46<02:46, 5.46it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

22%|██████████| 261/1167 [00:46<02:41, 5.61it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

22%|██████████| 262/1167 [00:46<02:40, 5.63it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

23%|██████████| 263/1167 [00:46<02:38, 5.69it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

23%|██████████ | 264/1167 [00:46<02:38, 5.70it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

23%|██████████ | 265/1167 [00:47<02:45, 5.44it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

23%|██████████| 266/1167 [00:47<02:43, 5.51it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

23%|██████████| 267/1167 [00:47<02:38, 5.67it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

23%|██████████| 268/1167 [00:47<02:37, 5.72it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

23%|██████████| 269/1167 [00:47<02:44, 5.45it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

23%|██████████| 270/1167 [00:47<02:43, 5.47it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

23%|██████████| 271/1167 [00:48<02:39, 5.63it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

23%|██████████ | 272/1167 [00:48<02:38, 5.64it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

23%|██████████ | 273/1167 [00:48<02:37, 5.68it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

23%|██████████| 274/1167 [00:48<02:38, 5.64it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

24%|██████████| 275/1167 [00:48<02:36, 5.71it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

24%|██████████| 276/1167 [00:49<02:36, 5.68it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

24%|██████████| 277/1167 [00:49<02:42, 5.49it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

24%|██████████| 278/1167 [00:49<02:42, 5.48it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

24%|██████████| 279/1167 [00:49<02:44, 5.40it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

24%|██████████ | 280/1167 [00:49<02:38, 5.59it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

24%|██████████ | 281/1167 [00:49<02:39, 5.54it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

24%|██████████| 282/1167 [00:50<02:42, 5.43it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

24%|██████████| 283/1167 [00:50<02:38, 5.59it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

24%|██████████| 284/1167 [00:50<02:36, 5.64it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

24%|██████████| 285/1167 [00:50<02:36, 5.64it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

25%|██████████| 286/1167 [00:50<02:36, 5.65it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

25%|██████████| 287/1167 [00:50<02:37, 5.57it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

25%|██████████ | 288/1167 [00:51<02:38, 5.55it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformersLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

25%|██████████ | 289/1167 [00:51<02:41, 5.44it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

25%|██████████| 290/1167 [00:51<02:37, 5.55it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

25%|██████████| 291/1167 [00:51<02:35, 5.62it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

25%|██████████| 292/1167 [00:51<02:37, 5.57it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

25%|██████████| 293/1167 [00:52<02:36, 5.60it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

25%|██████████| 294/1167 [00:52<02:37, 5.55it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

25%|██████████| 295/1167 [00:52<02:33, 5.66it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

25%|██████████ | 296/1167 [00:52<02:35, 5.60it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

25%|██████████ | 297/1167 [00:52<02:40, 5.43it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

26%|██████████| 298/1167 [00:52<02:41, 5.37it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

26%|██████████| 299/1167 [00:53<02:39, 5.45it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

26%|██████████| 300/1167 [00:53<02:35, 5.56it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

26%|██████████| 301/1167 [00:53<02:33, 5.62it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

26%|██████████| 302/1167 [00:53<02:32, 5.69it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

26%|██████████| 303/1167 [00:53<02:31, 5.71it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

26%|██████████ | 304/1167 [00:54<02:33, 5.61it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformersLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

26%|██████████ | 305/1167 [00:54<02:32, 5.65it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

26%|██████████| 306/1167 [00:54<02:34, 5.57it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

26%|██████████| 307/1167 [00:54<02:37, 5.46it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

26%|██████████| 308/1167 [00:54<02:34, 5.55it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

26%|██████████| 309/1167 [00:54<02:34, 5.55it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

27%|██████████| 310/1167 [00:55<02:37, 5.44it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

27%|██████████| 311/1167 [00:55<02:35, 5.49it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

27%|██████████ | 312/1167 [00:55<02:35, 5.51it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

27%|██████████ | 313/1167 [00:55<02:32, 5.59it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

27%|██████████| 314/1167 [00:55<02:41, 5.30it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

27%|██████████| 315/1167 [00:56<02:42, 5.25it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

27%|██████████| 316/1167 [00:56<02:38, 5.38it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

27%|██████████| 317/1167 [00:56<02:31, 5.60it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

27%|██████████| 318/1167 [00:56<02:31, 5.62it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

27%|██████████| 319/1167 [00:56<02:28, 5.71it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

27%|██████████ | 320/1167 [00:56<02:32, 5.57it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

28%|██████████ | 321/1167 [00:57<02:29, 5.67it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

28%|██████████| 322/1167 [00:57<02:27, 5.71it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

28%|██████████| 323/1167 [00:57<02:32, 5.54it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

28%|██████████| 324/1167 [00:57<02:30, 5.60it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

28%|██████████| 325/1167 [00:57<02:29, 5.63it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

28%|██████████| 326/1167 [00:58<02:34, 5.44it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

28%|██████████| 327/1167 [00:58<02:31, 5.56it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

28%|██████████ | 328/1167 [00:58<02:35, 5.40it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

28%|██████████ | 329/1167 [00:58<02:33, 5.45it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

28%|██████████| 330/1167 [00:58<02:37, 5.31it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

28%|██████████| 331/1167 [00:58<02:34, 5.42it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

28%|██████████| 332/1167 [00:59<02:31, 5.50it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

29%|██████████| 333/1167 [00:59<02:36, 5.32it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

29%|██████████| 334/1167 [00:59<02:32, 5.46it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

29%|██████████| 335/1167 [00:59<02:30, 5.54it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

29%|██████████ | 336/1167 [00:59<02:37, 5.27it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

29%|██████████ | 337/1167 [01:00<02:33, 5.39it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

29%|██████████| 338/1167 [01:00<02:34, 5.38it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

29%|██████████| 339/1167 [01:00<02:31, 5.47it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

29%|██████████| 340/1167 [01:00<02:28, 5.56it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

29%|██████████| 341/1167 [01:00<02:36, 5.29it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

29%|██████████| 342/1167 [01:00<02:32, 5.40it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

29%|██████████| 343/1167 [01:01<02:31, 5.43it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

29%|██████████ | 344/1167 [01:01<02:32, 5.41it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

30%|██████████ | 345/1167 [01:01<02:29, 5.51it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

30%|██████████| 346/1167 [01:01<02:26, 5.59it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

30%|██████████| 347/1167 [01:01<02:25, 5.62it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

30%|██████████| 348/1167 [01:02<02:28, 5.50it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

30%|██████████| 349/1167 [01:02<02:27, 5.53it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

30%|██████████| 350/1167 [01:02<02:27, 5.53it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

30%|██████████| 351/1167 [01:02<02:25, 5.60it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

30%|██████████ | 352/1167 [01:02<02:23, 5.68it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

30%|██████████ | 353/1167 [01:02<02:24, 5.63it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

30%|██████████| 354/1167 [01:03<02:25, 5.60it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

30%|██████████| 355/1167 [01:03<02:26, 5.55it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

31%|██████████| 356/1167 [01:03<02:35, 5.23it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

31%|██████████| 357/1167 [01:03<02:31, 5.36it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

31%|██████████| 358/1167 [01:03<02:28, 5.46it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

31%|██████████| 359/1167 [01:04<02:34, 5.23it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

31%|██████████ | 360/1167 [01:04<02:33, 5.27it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

31%|██████████ | 361/1167 [01:04<02:36, 5.14it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

31%|██████████| 362/1167 [01:04<02:33, 5.23it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

31%|██████████| 363/1167 [01:04<02:29, 5.39it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

31%|██████████| 364/1167 [01:05<02:40, 4.99it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

31%|██████████| 365/1167 [01:05<02:36, 5.13it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

31%|██████████| 366/1167 [01:05<02:42, 4.94it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

31%|██████████| 367/1167 [01:05<02:46, 4.81it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

32%|██████████ | 368/1167 [01:05<02:44, 4.86it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

32%|██████████ | 369/1167 [01:06<02:39, 5.00it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

32%|██████████| 370/1167 [01:06<02:33, 5.21it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

32%|██████████| 371/1167 [01:06<02:28, 5.34it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

32%|██████████| 372/1167 [01:06<02:25, 5.46it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

32%|██████████| 373/1167 [01:06<02:24, 5.51it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

32%|██████████| 374/1167 [01:06<02:22, 5.55it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

32%|██████████| 375/1167 [01:07<02:20, 5.63it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

32%|██████████ | 376/1167 [01:07<02:20, 5.62it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformersLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

32%|██████████ | 377/1167 [01:07<02:19, 5.66it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

32%|██████████| 378/1167 [01:07<02:20, 5.62it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

32%|██████████| 379/1167 [01:07<02:17, 5.73it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

33%|██████████| 380/1167 [01:08<02:29, 5.25it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

33%|██████████| 381/1167 [01:08<02:28, 5.31it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

33%|██████████| 382/1167 [01:08<02:26, 5.37it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

33%|██████████| 383/1167 [01:08<02:30, 5.22it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

33%|██████████ | 384/1167 [01:08<02:25, 5.38it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

33%|██████████ | 385/1167 [01:08<02:22, 5.50it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

33%|██████████| 386/1167 [01:09<02:28, 5.27it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

33%|██████████| 387/1167 [01:09<02:27, 5.29it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

33%|██████████| 388/1167 [01:09<02:22, 5.46it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

33%|██████████| 389/1167 [01:09<02:23, 5.42it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

33%|██████████| 390/1167 [01:09<02:31, 5.14it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

34%|██████████| 391/1167 [01:10<02:29, 5.20it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

34%|██████████ | 392/1167 [01:10<02:26, 5.30it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

34%|██████████ | 393/1167 [01:10<02:24, 5.36it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

34%|██████████| 394/1167 [01:10<02:24, 5.35it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

34%|██████████| 395/1167 [01:10<02:22, 5.40it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

34%|██████████| 396/1167 [01:11<02:24, 5.34it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

34%|██████████| 397/1167 [01:11<02:20, 5.49it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

34%|██████████| 398/1167 [01:11<02:16, 5.61it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

34%|██████████| 399/1167 [01:11<02:17, 5.59it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

34%|██████████ | 400/1167 [01:11<02:13, 5.75it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

34%|██████████ | 401/1167 [01:11<02:18, 5.53it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

34%|██████████| 402/1167 [01:12<02:32, 5.01it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

35%|██████████| 403/1167 [01:12<02:28, 5.14it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxFForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

35%|██████████| 404/1167 [01:12<02:27, 5.16it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxFForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

35%|██████████| 405/1167 [01:12<02:22, 5.35it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

35%|██████████| 406/1167 [01:12<02:19, 5.47it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

35%|██████████| 407/1167 [01:13<02:18, 5.48it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

35%|██████████ | 408/1167 [01:13<02:18, 5.47it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformersLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

35%|██████████ | 409/1167 [01:13<02:22, 5.33it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

35%|██████████| 410/1167 [01:13<02:23, 5.26it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

35%|██████████| 411/1167 [01:13<02:21, 5.34it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

35%|██████████| 412/1167 [01:14<02:19, 5.41it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

35%|██████████| 413/1167 [01:14<02:19, 5.40it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

35%|██████████| 414/1167 [01:14<02:33, 4.92it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

36%|██████████| 415/1167 [01:14<02:27, 5.10it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

36%|██████████ | 416/1167 [01:14<02:24, 5.20it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

36%|██████████ | 417/1167 [01:14<02:19, 5.37it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

36%|██████████| 418/1167 [01:15<02:20, 5.35it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

36%|██████████| 419/1167 [01:15<02:24, 5.19it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

36%|██████████| 420/1167 [01:15<02:20, 5.33it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

36%|██████████| 421/1167 [01:15<02:17, 5.44it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

36%|██████████| 422/1167 [01:15<02:17, 5.42it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

36%|██████████| 423/1167 [01:16<02:20, 5.28it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

36%|██████████ | 424/1167 [01:16<02:27, 5.04it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformersLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

36%|██████████ | 425/1167 [01:16<02:30, 4.93it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

37%|██████████| 426/1167 [01:16<02:23, 5.15it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

37%|██████████| 427/1167 [01:16<02:18, 5.32it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

37%|██████████| 428/1167 [01:17<02:18, 5.34it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

37%|██████████| 429/1167 [01:17<02:21, 5.23it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

37%|██████████| 430/1167 [01:17<02:17, 5.36it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

37%|██████████| 431/1167 [01:17<02:15, 5.43it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

37%|██████████ | 432/1167 [01:17<02:14, 5.45it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

37%|██████████ | 433/1167 [01:18<02:14, 5.46it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

37%|██████████| 434/1167 [01:18<02:17, 5.33it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

37%|██████████| 435/1167 [01:18<02:22, 5.12it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

37%|██████████| 436/1167 [01:18<02:22, 5.15it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

37%|██████████| 437/1167 [01:18<02:16, 5.36it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

38%|██████████| 438/1167 [01:18<02:13, 5.46it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

38%|██████████| 439/1167 [01:19<02:11, 5.52it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

38%|██████████ | 440/1167 [01:19<02:11, 5.52it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

38%|██████████ | 441/1167 [01:19<02:09, 5.59it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

38%|██████████| 442/1167 [01:19<02:13, 5.42it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

38%|██████████| 443/1167 [01:19<02:12, 5.48it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

38%|██████████| 444/1167 [01:20<02:11, 5.50it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

38%|██████████| 445/1167 [01:20<02:22, 5.08it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

38%|██████████| 446/1167 [01:20<02:18, 5.21it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

38%|██████████| 447/1167 [01:20<02:14, 5.34it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

38%|██████████ | 448/1167 [01:20<02:17, 5.22it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

38%|██████████ | 449/1167 [01:21<02:18, 5.18it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

39%|██████████| 450/1167 [01:21<02:15, 5.30it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

39%|██████████| 451/1167 [01:21<02:15, 5.30it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

39%|██████████| 452/1167 [01:21<02:14, 5.32it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

39%|██████████| 453/1167 [01:21<02:16, 5.21it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

39%|██████████| 454/1167 [01:21<02:17, 5.18it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

39%|██████████| 455/1167 [01:22<02:14, 5.30it/s]Device set to use cuda:0
The model 'PefitModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

39%|██████████ | 456/1167 [01:22<02:15, 5.23it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

39%|██████████ | 457/1167 [01:22<02:13, 5.30it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

39%|██████████| 458/1167 [01:22<02:17, 5.15it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

39%|██████████| 459/1167 [01:22<02:15, 5.24it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

39%|██████████| 460/1167 [01:23<02:12, 5.34it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

40%|██████████| 461/1167 [01:23<02:11, 5.39it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

40%|██████████| 462/1167 [01:23<02:09, 5.45it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

40%|██████████| 463/1167 [01:23<02:18, 5.08it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

40%|██████████ | 464/1167 [01:23<02:16, 5.17it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

40%|██████████ | 465/1167 [01:24<02:12, 5.28it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

40%|██████████| 466/1167 [01:24<02:11, 5.32it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

40%|██████████| 467/1167 [01:24<02:13, 5.22it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

40%|██████████| 468/1167 [01:24<02:16, 5.12it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

40%|██████████| 469/1167 [01:24<02:12, 5.26it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

40%|██████████| 470/1167 [01:25<02:10, 5.32it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

40%|██████████| 471/1167 [01:25<02:09, 5.37it/s]Device set to use cuda:0
The model 'PefitModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

40%|██████████ | 472/1167 [01:25<02:15, 5.11it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformersLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

41%|██████████ | 473/1167 [01:25<02:18, 5.02it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

41%|██████████| 474/1167 [01:25<02:13, 5.19it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

41%|██████████| 475/1167 [01:25<02:10, 5.30it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

41%|██████████| 476/1167 [01:26<02:06, 5.48it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

41%|██████████| 477/1167 [01:26<02:09, 5.33it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

41%|██████████| 478/1167 [01:26<02:07, 5.42it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

41%|██████████| 479/1167 [01:26<02:13, 5.16it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

41%|██████████ | 480/1167 [01:26<02:08, 5.33it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

41%|██████████ | 481/1167 [01:27<02:05, 5.46it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

41%|██████████| 482/1167 [01:27<02:16, 5.03it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

41%|██████████| 483/1167 [01:27<02:13, 5.11it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

41%|██████████| 484/1167 [01:27<02:12, 5.17it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

42%|██████████| 485/1167 [01:27<02:09, 5.27it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

42%|██████████| 486/1167 [01:28<02:06, 5.37it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

42%|██████████| 487/1167 [01:28<02:07, 5.32it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

42%|██████████ | 488/1167 [01:28<02:13, 5.08it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

42%|██████████ | 489/1167 [01:28<02:09, 5.25it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

42%|██████████| 490/1167 [01:28<02:11, 5.15it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

42%|██████████| 491/1167 [01:29<02:16, 4.94it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxFForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmForCausalLM', 'Olm2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

42%|██████████| 492/1167 [01:29<02:16, 4.96it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxFForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmForCausalLM', 'Olm2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

42%|██████████| 493/1167 [01:29<02:12, 5.10it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

42%|██████████| 494/1167 [01:29<02:07, 5.29it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

42%|██████████| 495/1167 [01:29<02:03, 5.45it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

43%|██████████ | 496/1167 [01:29<02:05, 5.34it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformersLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

43%|██████████ | 497/1167 [01:30<02:04, 5.37it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

43%|██████████| 498/1167 [01:30<02:05, 5.34it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

43%|██████████| 499/1167 [01:30<02:08, 5.21it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

43%|██████████| 500/1167 [01:30<02:03, 5.41it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

43%|██████████| 501/1167 [01:30<02:00, 5.53it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

43%|██████████| 502/1167 [01:31<01:58, 5.61it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

43%|██████████| 503/1167 [01:31<01:58, 5.6lit/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

43%|██████████ | 504/1167 [01:31<02:02, 5.41it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

43%|██████████ | 505/1167 [01:31<02:00, 5.49it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

43%|██████████| 506/1167 [01:31<02:00, 5.50it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

43%|██████████| 507/1167 [01:32<02:03, 5.32it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

44%|██████████| 508/1167 [01:32<02:01, 5.43it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

44%|██████████| 509/1167 [01:32<02:01, 5.40it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

44%|██████████| 510/1167 [01:32<02:02, 5.38it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

44%|██████████| 511/1167 [01:32<02:04, 5.28it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

44%|██████████ | 512/1167 [01:32<02:05, 5.20it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

44%|██████████ | 513/1167 [01:33<02:03, 5.28it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

44%|██████████| 514/1167 [01:33<02:01, 5.38it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

44%|██████████| 515/1167 [01:33<02:09, 5.04it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

44%|██████████| 516/1167 [01:33<02:06, 5.15it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

44%|██████████| 517/1167 [01:33<02:01, 5.34it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

44%|██████████| 518/1167 [01:34<02:09, 5.02it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

44%|██████████| 519/1167 [01:34<02:05, 5.16it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

45%|██████████ | 520/1167 [01:34<02:02, 5.26it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformersLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

45%|██████████ | 521/1167 [01:34<02:05, 5.14it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

45%|██████████| 522/1167 [01:34<02:02, 5.27it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

45%|██████████| 523/1167 [01:35<02:00, 5.36it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

45%|██████████| 524/1167 [01:35<02:01, 5.28it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

45%|██████████| 525/1167 [01:35<02:10, 4.91it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

45%|██████████| 526/1167 [01:35<02:08, 4.99it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

45%|██████████| 527/1167 [01:35<02:02, 5.23it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

45%|██████████ | 528/1167 [01:36<01:58, 5.38it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

45%|██████████ | 529/1167 [01:36<01:58, 5.38it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

45%|██████████| 530/1167 [01:36<02:00, 5.31it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

46%|██████████| 531/1167 [01:36<02:03, 5.16it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

46%|██████████| 532/1167 [01:36<02:09, 4.89it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

46%|██████████| 533/1167 [01:37<02:05, 5.04it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

46%|██████████| 534/1167 [01:37<02:03, 5.12it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

46%|██████████| 535/1167 [01:37<02:07, 4.97it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

46%|██████████ | 536/1167 [01:37<02:01, 5.20it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformersLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

46%|██████████ | 537/1167 [01:37<01:58, 5.33it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

46%|██████████| 538/1167 [01:37<01:54, 5.47it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

46%|██████████| 539/1167 [01:38<02:02, 5.12it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

46%|██████████| 540/1167 [01:38<02:04, 5.05it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

46%|██████████| 541/1167 [01:38<02:04, 5.01it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

46%|██████████| 542/1167 [01:38<01:59, 5.25it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

47%|██████████| 543/1167 [01:38<02:04, 5.03it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

47%|██████████ | 544/1167 [01:39<01:58, 5.25it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

47%|██████████ | 545/1167 [01:39<02:03, 5.05it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

47%|██████████| 546/1167 [01:39<01:59, 5.19it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

47%|██████████| 547/1167 [01:39<01:56, 5.33it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

47%|██████████| 548/1167 [01:39<01:52, 5.51it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

47%|██████████| 549/1167 [01:40<01:51, 5.57it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

47%|██████████| 550/1167 [01:40<01:52, 5.46it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

47%|██████████| 551/1167 [01:40<01:55, 5.33it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

47%|██████████ | 552/1167 [01:40<01:54, 5.35it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

47%|██████████ | 553/1167 [01:40<01:52, 5.46it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

47%|██████████| 554/1167 [01:40<01:52, 5.44it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

48%|██████████| 555/1167 [01:41<01:52, 5.43it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

48%|██████████| 556/1167 [01:41<01:52, 5.45it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

48%|██████████| 557/1167 [01:41<01:54, 5.34it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

48%|██████████| 558/1167 [01:41<01:57, 5.20it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

48%|██████████| 559/1167 [01:41<01:58, 5.15it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

48%|██████████ | 560/1167 [01:42<01:54, 5.32it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

48%|██████████ | 561/1167 [01:42<02:02, 4.95it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

48%|██████████| 562/1167 [01:42<02:04, 4.88it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

48%|██████████| 563/1167 [01:42<02:02, 4.92it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

48%|██████████| 564/1167 [01:42<02:01, 4.98it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

48%|██████████| 565/1167 [01:43<01:58, 5.08it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

49%|██████████| 566/1167 [01:43<01:55, 5.20it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

49%|██████████| 567/1167 [01:43<01:52, 5.32it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

49%|██████████ | 568/1167 [01:43<01:53, 5.26it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

49%|██████████ | 569/1167 [01:43<01:51, 5.38it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

49%|██████████| 570/1167 [01:44<01:50, 5.42it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

49%|██████████| 571/1167 [01:44<01:53, 5.24it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

49%|██████████| 572/1167 [01:44<01:57, 5.07it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

49%|██████████| 573/1167 [01:44<01:55, 5.15it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

49%|██████████| 574/1167 [01:44<01:53, 5.21it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

49%|██████████| 575/1167 [01:45<01:52, 5.28it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

49%|██████████ | 576/1167 [01:45<01:54, 5.16it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

49%|██████████ | 577/1167 [01:45<01:55, 5.12it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

50%|██████████| 578/1167 [01:45<02:02, 4.80it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

50%|██████████| 579/1167 [01:45<02:03, 4.78it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

50%|██████████| 580/1167 [01:46<01:57, 5.00it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

50%|██████████| 581/1167 [01:46<01:58, 4.95it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

50%|██████████| 582/1167 [01:46<02:01, 4.82it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

50%|██████████| 583/1167 [01:46<01:57, 4.96it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

50%|██████████ | 584/1167 [01:46<01:57, 4.97it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

50%|██████████ | 585/1167 [01:47<01:54, 5.09it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

50%|██████████| 586/1167 [01:47<01:54, 5.08it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

50%|██████████| 587/1167 [01:47<01:53, 5.13it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

50%|██████████| 588/1167 [01:47<01:51, 5.21it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

50%|██████████| 589/1167 [01:47<01:53, 5.11it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

51%|██████████| 590/1167 [01:48<01:51, 5.18it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

51%|██████████| 591/1167 [01:48<01:49, 5.27it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

51%|██████████ | 592/1167 [01:48<01:48, 5.31it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

51%|██████████ | 593/1167 [01:48<01:49, 5.25it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

51%|██████████| 594/1167 [01:48<01:46, 5.39it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

51%|██████████| 595/1167 [01:48<01:51, 5.15it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

51%|██████████| 596/1167 [01:49<01:47, 5.31lit/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

51%|██████████| 597/1167 [01:49<01:48, 5.24it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

51%|██████████| 598/1167 [01:49<01:54, 4.96it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

51%|██████████| 599/1167 [01:49<01:54, 4.95it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

51%|██████████ | 600/1167 [01:49<01:51, 5.11it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

51%|██████████ | 601/1167 [01:50<01:47, 5.25it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

52%|██████████| 602/1167 [01:50<01:49, 5.17it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

52%|██████████| 603/1167 [01:50<01:50, 5.08it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

52%|██████████| 604/1167 [01:50<01:55, 4.88it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

52%|██████████| 605/1167 [01:50<01:50, 5.07it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

52%|██████████| 606/1167 [01:51<01:48, 5.19it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

52%|██████████| 607/1167 [01:51<01:46, 5.24it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

52%|██████████ | 608/1167 [01:51<01:43, 5.39it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

52%|██████████ | 609/1167 [01:51<01:48, 5.16it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

52%|██████████| 610/1167 [01:51<01:49, 5.08it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

52%|██████████| 611/1167 [01:52<01:46, 5.24it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformersLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

52%|██████████| 612/1167 [01:52<01:49, 5.08it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

53%|██████████| 613/1167 [01:52<01:47, 5.16it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

53%|██████████| 614/1167 [01:52<01:47, 5.14it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

53%|██████████| 615/1167 [01:52<01:50, 4.98it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

53%|██████████ | 616/1167 [01:53<01:49, 5.03it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

53%|██████████ | 617/1167 [01:53<01:45, 5.23it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

53%|██████████| 618/1167 [01:53<01:45, 5.22it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

53%|██████████| 619/1167 [01:53<01:49, 5.00it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

53%|██████████| 620/1167 [01:53<01:46, 5.15it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

53%|██████████| 621/1167 [01:54<01:45, 5.18it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

53%|██████████| 622/1167 [01:54<01:47, 5.07it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

53%|██████████| 623/1167 [01:54<01:47, 5.06it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

53%|██████████ | 624/1167 [01:54<01:49, 4.95it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

54%|██████████ | 625/1167 [01:54<01:49, 4.97it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

54%|██████████| 626/1167 [01:55<01:48, 4.99it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

54%|██████████| 627/1167 [01:55<01:44, 5.15it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

54%|██████████| 628/1167 [01:55<01:49, 4.90it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

54%|██████████| 629/1167 [01:55<01:49, 4.93it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

54%|██████████| 630/1167 [01:55<01:45, 5.09it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

54%|██████████| 631/1167 [01:56<01:43, 5.19it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

54%|██████████ | 632/1167 [01:56<01:39, 5.38it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

54%|██████████ | 633/1167 [01:56<01:43, 5.15it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

54%|██████████| 634/1167 [01:56<01:45, 5.05it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

54%|██████████| 635/1167 [01:56<01:44, 5.10it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

54%|██████████| 636/1167 [01:56<01:41, 5.22it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

55%|██████████| 637/1167 [01:57<01:39, 5.32it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

55%|██████████| 638/1167 [01:57<01:40, 5.27it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

55%|██████████| 639/1167 [01:57<01:41, 5.21it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

55%|██████████ | 640/1167 [01:57<01:38, 5.35it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

55%|██████████ | 641/1167 [01:57<01:41, 5.19it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

55%|██████████| 642/1167 [01:58<01:42, 5.11it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

55%|██████████| 643/1167 [01:58<01:46, 4.94it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

55%|██████████| 644/1167 [01:58<01:44, 5.00it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

55%|██████████| 645/1167 [01:58<01:47, 4.84it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

55%|██████████| 646/1167 [01:58<01:44, 4.96it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

55%|██████████| 647/1167 [01:59<01:46, 4.87it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

56%|██████████ | 648/1167 [01:59<01:44, 4.94it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

56%|██████████ | 649/1167 [01:59<01:45, 4.90it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

56%|██████████| 650/1167 [01:59<01:50, 4.69it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

56%|██████████| 651/1167 [02:00<01:51, 4.62it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

56%|██████████| 652/1167 [02:00<01:46, 4.85it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

56%|██████████| 653/1167 [02:00<01:48, 4.73it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

56%|██████████| 654/1167 [02:00<01:45, 4.87it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

56%|██████████| 655/1167 [02:00<01:42, 5.0lit/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

56%|██████████ | 656/1167 [02:01<01:42, 5.00it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

56%|██████████ | 657/1167 [02:01<01:43, 4.92it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

56%|██████████| 658/1167 [02:01<01:41, 5.00it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

56%|██████████| 659/1167 [02:01<01:41, 5.01it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

57%|██████████| 660/1167 [02:01<01:41, 5.01it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

57%|██████████| 661/1167 [02:02<01:38, 5.12it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

57%|██████████| 662/1167 [02:02<01:35, 5.29it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

57%|██████████| 663/1167 [02:02<01:38, 5.10it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

57%|██████████ | 664/1167 [02:02<01:38, 5.10it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformersLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

57%|██████████ | 665/1167 [02:02<01:39, 5.05it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

57%|██████████| 666/1167 [02:03<01:41, 4.94it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

57%|██████████| 667/1167 [02:03<01:41, 4.93it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

57%|██████████| 668/1167 [02:03<01:46, 4.67it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

57%|██████████| 669/1167 [02:03<01:43, 4.81it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

57%|██████████| 670/1167 [02:03<01:44, 4.78it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

57%|██████████| 671/1167 [02:04<01:40, 4.91lit/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

58%|██████████ | 672/1167 [02:04<01:38, 5.02it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

58%|██████████ | 673/1167 [02:04<01:36, 5.13it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

58%|██████████| 674/1167 [02:04<01:38, 5.01it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

58%|██████████| 675/1167 [02:04<01:35, 5.16it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

58%|██████████| 676/1167 [02:05<01:38, 5.00it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

58%|██████████| 677/1167 [02:05<01:37, 5.02it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

58%|██████████| 678/1167 [02:05<01:44, 4.68it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

58%|██████████| 679/1167 [02:05<01:40, 4.86it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

58%|██████████ | 680/1167 [02:05<01:40, 4.83it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

58%|██████████ | 681/1167 [02:06<01:37, 5.01it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

58%|██████████| 682/1167 [02:06<01:38, 4.93it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

59%|██████████| 683/1167 [02:06<01:41, 4.78it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

59%|██████████| 684/1167 [02:06<01:38, 4.90it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

59%|██████████| 685/1167 [02:06<01:34, 5.09it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

59%|██████████| 686/1167 [02:07<01:32, 5.20it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

59%|██████████| 687/1167 [02:07<01:30, 5.30it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

59%|██████████ | 688/1167 [02:07<01:31, 5.25it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

59%|██████████ | 689/1167 [02:07<01:29, 5.32it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

59%|██████████| 690/1167 [02:07<01:35, 5.00it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

59%|██████████| 691/1167 [02:08<01:33, 5.07it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

59%|██████████| 692/1167 [02:08<01:36, 4.95it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

59%|██████████| 693/1167 [02:08<01:37, 4.84it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

59%|██████████| 694/1167 [02:08<01:37, 4.88it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

60%|██████████| 695/1167 [02:08<01:35, 4.95it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

60%|██████████ | 696/1167 [02:09<01:32, 5.08it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

60%|██████████ | 697/1167 [02:09<01:36, 4.89it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

60%|██████████| 698/1167 [02:09<01:33, 5.03it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

60%|██████████| 699/1167 [02:09<01:34, 4.98it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformersLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

60%|██████████| 700/1167 [02:09<01:36, 4.86it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

60%|██████████| 701/1167 [02:10<01:32, 5.04it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

60%|██████████| 702/1167 [02:10<01:33, 4.95it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

60%|██████████| 703/1167 [02:10<01:36, 4.83it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

60%|██████████ | 704/1167 [02:10<01:37, 4.77it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

60%|██████████ | 705/1167 [02:10<01:37, 4.72it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

60%|██████████ | 706/1167 [02:11<01:36, 4.79it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

61%|██████████ | 707/1167 [02:11<01:34, 4.89it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxFForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

61%|██████████ | 708/1167 [02:11<01:31, 5.02it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxFForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

61%|██████████| 709/1167 [02:11<01:29, 5.14it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

61%|██████████| 710/1167 [02:11<01:29, 5.09it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

61%|██████████| 711/1167 [02:12<01:26, 5.30it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

61%|██████████ | 712/1167 [02:12<01:24, 5.37it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

61%|██████████ | 713/1167 [02:12<01:24, 5.36it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

61%|██████████| 714/1167 [02:12<01:29, 5.07it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

61%|██████████| 715/1167 [02:12<01:32, 4.90it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

61%|██████████| 716/1167 [02:13<01:32, 4.88it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

61%|██████████| 717/1167 [02:13<01:36, 4.68it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

62%|██████████| 718/1167 [02:13<01:36, 4.66it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

62%|██████████ | 719/1167 [02:13<01:32, 4.83it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

62%|██████████ | 720/1167 [02:13<01:33, 4.79it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

62%|██████████ | 721/1167 [02:14<01:32, 4.81it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

62%|██████████| 722/1167 [02:14<01:31, 4.86it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

62%|██████████| 723/1167 [02:14<01:32, 4.83it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

62%|██████████| 724/1167 [02:14<01:29, 4.96it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

62%|██████████| 725/1167 [02:14<01:30, 4.87it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

62%|██████████| 726/1167 [02:15<01:31, 4.82it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

62%|██████████ | 727/1167 [02:15<01:28, 5.00it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

62%|██████████ | 728/1167 [02:15<01:32, 4.77it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

62%|██████████ | 729/1167 [02:15<01:30, 4.84it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

63%|██████████| 730/1167 [02:16<01:34, 4.62it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

63%|██████████| 731/1167 [02:16<01:31, 4.76it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

63%|██████████ | 732/1167 [02:16<01:28, 4.92it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

63%|██████████ | 733/1167 [02:16<01:36, 4.52it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

63%|██████████ | 734/1167 [02:16<01:30, 4.76it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

63%|██████████ | 735/1167 [02:17<01:29, 4.82it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

63%|██████████ | 736/1167 [02:17<01:26, 5.00it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

63%|██████████ | 737/1167 [02:17<01:27, 4.92it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

63%|██████████| 738/1167 [02:17<01:25, 5.05it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

63%|██████████| 739/1167 [02:17<01:26, 4.96it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

63%|██████████| 740/1167 [02:18<01:22, 5.18it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

63%|██████████| 741/1167 [02:18<01:24, 5.05it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

64%|██████████| 742/1167 [02:18<01:22, 5.13it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

64%|██████████ | 743/1167 [02:18<01:24, 5.00it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

64%|██████████ | 744/1167 [02:18<01:24, 4.99it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

64%|██████████ | 745/1167 [02:19<01:26, 4.89it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

64%|██████████ | 746/1167 [02:19<01:26, 4.89it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

64%|██████████ | 747/1167 [02:19<01:26, 4.88it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

64%|██████████ | 748/1167 [02:19<01:22, 5.08it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

64%|██████████ | 749/1167 [02:19<01:23, 4.98it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

64%|██████████ | 750/1167 [02:20<01:20, 5.18it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

64%|██████████ | 751/1167 [02:20<01:22, 5.02it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

64%|██████████ | 752/1167 [02:20<01:20, 5.18it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

65%|██████████ | 753/1167 [02:20<01:18, 5.26it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

65%|██████████ | 754/1167 [02:20<01:22, 5.03it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

65%|██████████ | 755/1167 [02:20<01:19, 5.19it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

65%|██████████ | 756/1167 [02:21<01:20, 5.13it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

65%|██████████| 757/1167 [02:21<01:20, 5.12it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

65%|██████████| 758/1167 [02:21<01:21, 5.01it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

65%|██████████ | 759/1167 [02:21<01:20, 5.08it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

65%|██████████ | 760/1167 [02:21<01:22, 4.93it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

65%|██████████ | 761/1167 [02:22<01:21, 5.00it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

65%|██████████ | 762/1167 [02:22<01:22, 4.89it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

65%|██████████ | 763/1167 [02:22<01:21, 4.94it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

65%|██████████ | 764/1167 [02:22<01:23, 4.85it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

66%|██████████| 765/1167 [02:22<01:18, 5.09it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

66%|██████████| 766/1167 [02:23<01:22, 4.85it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

66%|██████████ | 767/1167 [02:23<01:19, 5.03it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

66%|██████████ | 768/1167 [02:23<01:21, 4.92it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

66%|██████████ | 769/1167 [02:23<01:19, 5.03it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

66%|██████████| 770/1167 [02:24<01:19, 4.99it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

66%|██████████| 771/1167 [02:24<01:19, 4.98it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

66%|██████████ | 772/1167 [02:24<01:19, 4.98it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

66%|██████████| 773/1167 [02:24<01:16, 5.16it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

66%|██████████| 774/1167 [02:24<01:22, 4.79it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

66%|██████████ | 775/1167 [02:25<01:27, 4.49it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

66%|██████████ | 776/1167 [02:25<01:25, 4.60it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

67%|██████████ | 777/1167 [02:25<01:21, 4.81it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

67%|██████████ | 778/1167 [02:25<01:23, 4.63it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

67%|██████████ | 779/1167 [02:25<01:20, 4.81it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

67%|██████████ | 780/1167 [02:26<01:19, 4.84it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

67%|██████████| 781/1167 [02:26<01:18, 4.92it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

67%|██████████| 782/1167 [02:26<01:18, 4.91it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

67%|██████████ | 783/1167 [02:26<01:17, 4.95it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

67%|██████████ | 784/1167 [02:26<01:18, 4.89it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

67%|██████████ | 785/1167 [02:27<01:17, 4.96it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

67%|██████████ | 786/1167 [02:27<01:13, 5.18it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

67%|██████████ | 787/1167 [02:27<01:15, 5.06it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

68%|██████████ | 788/1167 [02:27<01:17, 4.91lit/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

68%|██████████| 789/1167 [02:27<01:20, 4.72it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

68%|██████████| 790/1167 [02:28<01:18, 4.78it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

68%|██████████ | 791/1167 [02:28<01:18, 4.81lit/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

68%|██████████ | 792/1167 [02:28<01:14, 5.02it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

68%|██████████ | 793/1167 [02:28<01:17, 4.86it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

68%|██████████ | 794/1167 [02:28<01:13, 5.04it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

68%|██████████ | 795/1167 [02:29<01:17, 4.82it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

68%|██████████| 796/1167 [02:29<01:15, 4.94it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

68%|██████████| 797/1167 [02:29<01:16, 4.86it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

68%|██████████| 798/1167 [02:29<01:18, 4.71it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

68%|██████████ | 799/1167 [02:30<01:19, 4.60it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

69%|██████████ | 800/1167 [02:30<01:20, 4.56it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

69%|██████████ | 801/1167 [02:30<01:19, 4.58it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

69%|██████████ | 802/1167 [02:30<01:21, 4.50it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

69%|██████████ | 803/1167 [02:30<01:22, 4.39it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

69%|██████████ | 804/1167 [02:31<01:21, 4.47it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

69%|██████████| 805/1167 [02:31<01:15, 4.76it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

69%|██████████| 806/1167 [02:31<01:13, 4.90it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

69%|██████████ | 807/1167 [02:31<01:12, 4.97it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

69%|██████████ | 808/1167 [02:31<01:10, 5.07it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

69%|██████████ | 809/1167 [02:32<01:08, 5.19it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

69%|██████████ | 810/1167 [02:32<01:09, 5.10it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

69%|██████████ | 811/1167 [02:32<01:11, 4.95it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

70%|██████████| 812/1167 [02:32<01:13, 4.83it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

70%|██████████| 813/1167 [02:32<01:10, 5.00it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

70%|██████████| 814/1167 [02:33<01:10, 5.01it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

70%|██████████| 815/1167 [02:33<01:08, 5.12it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

70%|██████████ | 816/1167 [02:33<01:06, 5.28it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

70%|██████████ | 817/1167 [02:33<01:05, 5.34it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

70%|██████████| 818/1167 [02:33<01:09, 5.02it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

70%|██████████| 819/1167 [02:34<01:08, 5.10it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

70%|██████████| 820/1167 [02:34<01:08, 5.06it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

70%|██████████| 821/1167 [02:34<01:10, 4.94it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

70%|██████████| 822/1167 [02:34<01:09, 5.00it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

71%|██████████| 823/1167 [02:34<01:08, 5.04it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

71%|██████████ | 824/1167 [02:35<01:06, 5.18it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

71%|██████████ | 825/1167 [02:35<01:10, 4.86it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

71%|██████████| 826/1167 [02:35<01:09, 4.92it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

71%|██████████| 827/1167 [02:35<01:07, 5.01it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

71%|██████████ | 828/1167 [02:35<01:06, 5.07it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

71%|██████████| 829/1167 [02:36<01:08, 4.96it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

71%|██████████| 830/1167 [02:36<01:06, 5.03it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

71% |██████████| 831/1167 [02:36<01:06, 5.07it/s] Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

71%|██████████ | 832/1167 [02:36<01:07, 4.95it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

71%|██████████ | 833/1167 [02:36<01:05, 5.06it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

71%|██████████| 834/1167 [02:37<01:04, 5.18it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

72%|██████████| 835/1167 [02:37<01:06, 5.01it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

72%|██████████| 836/1167 [02:37<01:06, 4.99it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

72%|██████████| 837/1167 [02:37<01:04, 5.11it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

72%|██████████| 838/1167 [02:37<01:08, 4.83it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

72%|██████████| 839/1167 [02:38<01:09, 4.75it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

72%|██████████ | 840/1167 [02:38<01:07, 4.84it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

72%|██████████ | 841/1167 [02:38<01:08, 4.73it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

72%|██████████| 842/1167 [02:38<01:07, 4.82it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

72%|██████████| 843/1167 [02:38<01:09, 4.67it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

72%|██████████| 844/1167 [02:39<01:06, 4.86it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

72%|██████████| 845/1167 [02:39<01:07, 4.78it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

72%|██████████| 846/1167 [02:39<01:06, 4.85it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

73%|██████████| 847/1167 [02:39<01:07, 4.74it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

73%|██████████ | 848/1167 [02:40<01:11, 4.47it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

73%|██████████ | 849/1167 [02:40<01:10, 4.51it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

73%|██████████| 850/1167 [02:40<01:10, 4.49it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

73%|██████████| 851/1167 [02:40<01:10, 4.45it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

73%|██████████| 852/1167 [02:40<01:07, 4.66it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

73%|██████████| 853/1167 [02:41<01:05, 4.77it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

73%|██████████| 854/1167 [02:41<01:05, 4.81it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

73%|██████████| 855/1167 [02:41<01:03, 4.94it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

73%|██████████ | 856/1167 [02:41<01:03, 4.92it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformersLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

73%|██████████ | 857/1167 [02:41<01:02, 4.98it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

74%|██████████| 858/1167 [02:42<01:02, 4.96it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

74%|██████████| 859/1167 [02:42<01:02, 4.90it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

74%|██████████| 860/1167 [02:42<01:03, 4.86it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

74%|██████████| 861/1167 [02:42<01:03, 4.85it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

74%|██████████| 862/1167 [02:42<01:03, 4.82it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

74%|██████████| 863/1167 [02:43<01:00, 5.04it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

74%|██████████ | 864/1167 [02:43<00:58, 5.14it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

74%|██████████ | 865/1167 [02:43<01:00, 5.02it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

74%|██████████| 866/1167 [02:43<01:00, 4.94it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

74%|██████████| 867/1167 [02:43<01:01, 4.90it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

74%|██████████| 868/1167 [02:44<00:59, 4.99it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

74%|██████████| 869/1167 [02:44<00:59, 5.02it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

75%|██████████| 870/1167 [02:44<00:58, 5.05it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

75%|██████████| 871/1167 [02:44<01:02, 4.76it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

75%|██████████ | 872/1167 [02:44<01:00, 4.91it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

75%|██████████ | 873/1167 [02:45<00:58, 5.00it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

75%|██████████ | 874/1167 [02:45<00:57, 5.08it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

75%|██████████ | 875/1167 [02:45<01:00, 4.79it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

75%|██████████ | 876/1167 [02:45<01:01, 4.75it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

75%|██████████| 877/1167 [02:45<01:01, 4.70it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

75%|██████████| 878/1167 [02:46<00:59, 4.84it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

75%|██████████| 879/1167 [02:46<00:58, 4.91lit/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

75%|██████████ | 880/1167 [02:46<00:57, 5.01it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

75%|██████████ | 881/1167 [02:46<00:57, 4.99it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

76%|██████████| 882/1167 [02:46<00:57, 5.00it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

76%|██████████| 883/1167 [02:47<00:58, 4.89it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

76%|██████████| 884/1167 [02:47<00:57, 4.96it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

76%|██████████| 885/1167 [02:47<00:56, 4.96it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

76%|██████████| 886/1167 [02:47<00:55, 5.09it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

76%|██████████ | 887/1167 [02:47<00:55, 5.00it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

76%|██████████ | 888/1167 [02:48<00:54, 5.14it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

76%|██████████ | 889/1167 [02:48<00:56, 4.95it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

76%|██████████ | 890/1167 [02:48<00:55, 4.97it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

76%|██████████ | 891/1167 [02:48<00:54, 5.03it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

76%|██████████ | 892/1167 [02:48<00:58, 4.68it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

77%|██████████| 893/1167 [02:49<00:58, 4.69it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

77%|██████████| 894/1167 [02:49<00:58, 4.64it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

77% |██████████| 895/1167 [02:49<00:59, 4.57it/s] Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

77%|██████████ | 896/1167 [02:49<00:56, 4.76it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

77%|██████████ | 897/1167 [02:50<00:54, 4.94it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

77%|██████████ | 898/1167 [02:50<00:53, 5.02it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

77%|██████████ | 899/1167 [02:50<00:52, 5.08it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

77%|██████████ | 900/1167 [02:50<00:52, 5.11it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

77%|██████████| 901/1167 [02:50<00:54, 4.87it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

77%|██████████| 902/1167 [02:51<00:54, 4.84it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

77%|██████████| 903/1167 [02:51<00:53, 4.97it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

77%|██████████ | 904/1167 [02:51<00:52, 5.00it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

78%|██████████ | 905/1167 [02:51<00:53, 4.93it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

78%|██████████ | 906/1167 [02:51<00:54, 4.79it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

78%|██████████ | 907/1167 [02:52<00:53, 4.85it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

78%|██████████ | 908/1167 [02:52<00:55, 4.68it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

78%|██████████ | 909/1167 [02:52<00:54, 4.73it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

78%|██████████ | 910/1167 [02:52<00:56, 4.53it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

78% |██████████| 911/1167 [02:52<00:55, 4.63it/s] Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

78%|██████████ | 912/1167 [02:53<00:55, 4.61it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformersLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

78%|██████████ | 913/1167 [02:53<00:52, 4.84it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

78%|██████████ | 914/1167 [02:53<00:53, 4.73it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

78%|██████████ | 915/1167 [02:53<00:51, 4.86it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

78%|██████████ | 916/1167 [02:53<00:52, 4.80it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

79%|██████████ | 917/1167 [02:54<00:51, 4.83it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

79%|██████████ | 918/1167 [02:54<00:53, 4.66it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

79%|██████████ | 919/1167 [02:54<00:50, 4.89it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

79%|██████████ | 920/1167 [02:54<00:51, 4.81it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

79%|██████████ | 921/1167 [02:55<00:52, 4.73it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

79%|██████████ | 922/1167 [02:55<00:49, 4.92it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

79%|██████████ | 923/1167 [02:55<00:48, 5.04it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

79%|██████████ | 924/1167 [02:55<00:46, 5.21it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

79%|██████████ | 925/1167 [02:55<00:45, 5.29it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

79%|██████████ | 926/1167 [02:55<00:48, 4.95it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

79%|██████████ | 927/1167 [02:56<00:48, 4.96it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

80%|██████████ | 928/1167 [02:56<00:48, 4.98it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

80%|██████████ | 929/1167 [02:56<00:49, 4.78it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

80%|██████████ | 930/1167 [02:56<00:51, 4.58it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

80%|██████████ | 931/1167 [02:57<00:49, 4.76it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

80%|██████████ | 932/1167 [02:57<00:49, 4.79it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

80%|██████████ | 933/1167 [02:57<00:47, 4.93it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

80%|██████████ | 934/1167 [02:57<00:47, 4.95it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

80%|██████████ | 935/1167 [02:57<00:48, 4.77it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

80%|██████████ | 936/1167 [02:58<00:48, 4.76it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

80%|██████████ | 937/1167 [02:58<00:48, 4.79it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

80%|██████████ | 938/1167 [02:58<00:46, 4.91it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

80%|██████████ | 939/1167 [02:58<00:46, 4.86it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

81%|██████████ | 940/1167 [02:58<00:46, 4.90it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

81% |██████████| 941/1167 [02:59<00:46, 4.90it/s] Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

81% |██████████| 942/1167 [02:59<00:46, 4.84it/s] Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

81% |██████████| 943/1167 [02:59<00:44, 5.06it/s] Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

81%|██████████ | 944/1167 [02:59<00:42, 5.20it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

81%|██████████ | 945/1167 [02:59<00:45, 4.92it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

81%|██████████ | 946/1167 [03:00<00:43, 5.04it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

81%|██████████ | 947/1167 [03:00<00:44, 4.94it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

81%|██████████ | 948/1167 [03:00<00:43, 5.09it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

81% |██████████| 949/1167 [03:00<00:42, 5.10it/s] Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

81% |██████████| 950/1167 [03:00<00:43, 4.94it/s] Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

81% |██████████| 951/1167 [03:01<00:44, 4.83it/s] Device set to use cuda:0
The model 'PefitModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

82%|██████████ | 952/1167 [03:01<00:42, 5.04it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

82%|██████████ | 953/1167 [03:01<00:41, 5.15it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

82%|██████████ | 954/1167 [03:01<00:40, 5.22it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

82%|██████████ | 955/1167 [03:01<00:41, 5.15it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

82%|██████████ | 956/1167 [03:02<00:40, 5.20it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

82%|██████████ | 957/1167 [03:02<00:41, 5.08it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

82%|██████████ | 958/1167 [03:02<00:42, 4.90it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

82% |██████████| 959/1167 [03:02<00:41, 5.0lit/s] Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

82%|██████████ | 960/1167 [03:02<00:42, 4.92it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

82%|██████████ | 961/1167 [03:03<00:41, 4.99it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

82%|██████████ | 962/1167 [03:03<00:39, 5.15it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

83%|██████████ | 963/1167 [03:03<00:40, 5.02it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

83%|██████████ | 964/1167 [03:03<00:39, 5.08it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

83%|██████████ | 965/1167 [03:03<00:42, 4.77it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

83%|██████████ | 966/1167 [03:04<00:43, 4.60it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

83% |██████████| 967/1167 [03:04<00:41, 4.77it/s] Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

83%|██████████ | 968/1167 [03:04<00:43, 4.55it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

83%|██████████ | 969/1167 [03:04<00:42, 4.64it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

83%|██████████ | 970/1167 [03:04<00:43, 4.55it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

83%|██████████ | 971/1167 [03:05<00:41, 4.78it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

83%|██████████ | 972/1167 [03:05<00:41, 4.68it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

83%|██████████ | 973/1167 [03:05<00:40, 4.81it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

83%|██████████ | 974/1167 [03:05<00:44, 4.34it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

84%|██████████ | 975/1167 [03:06<00:43, 4.45it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

84%|██████████ | 976/1167 [03:06<00:43, 4.42it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

84%|██████████ | 977/1167 [03:06<00:40, 4.65it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

84%|██████████ | 978/1167 [03:06<00:40, 4.62it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

84%|██████████ | 979/1167 [03:06<00:40, 4.63it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

84%|██████████ | 980/1167 [03:07<00:38, 4.83it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

84%|██████████ | 981/1167 [03:07<00:37, 4.97it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

84%|██████████ | 982/1167 [03:07<00:37, 4.88it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

84%|██████████ | 983/1167 [03:07<00:37, 4.89it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

84%|██████████ | 984/1167 [03:07<00:37, 4.91it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

84%|██████████ | 985/1167 [03:08<00:35, 5.12it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

84%|██████████ | 986/1167 [03:08<00:34, 5.21it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

85%|██████████ | 987/1167 [03:08<00:34, 5.19it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

85%|██████████ | 988/1167 [03:08<00:33, 5.32it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

85%|██████████ | 989/1167 [03:08<00:35, 5.01it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

85%|██████████ | 990/1167 [03:09<00:35, 4.93it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

85%|██████████ | 991/1167 [03:09<00:35, 5.0lit/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

85%|██████████ | 992/1167 [03:09<00:33, 5.19it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

85%|██████████ | 993/1167 [03:09<00:33, 5.23it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

85%|██████████ | 994/1167 [03:09<00:33, 5.17it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

85%|██████████ | 995/1167 [03:10<00:34, 4.97it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

85%|██████████ | 996/1167 [03:10<00:33, 5.11lit/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

85%|██████████ | 997/1167 [03:10<00:34, 4.97it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

86%|██████████ | 998/1167 [03:10<00:33, 4.98it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

86%|██████████ | 999/1167 [03:10<00:33, 4.99it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

86%|██████████ | 1000/1167 [03:11<00:34, 4.87it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

86%|██████████ | 1001/1167 [03:11<00:35, 4.71it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

86%|██████████ | 1002/1167 [03:11<00:35, 4.71it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

86%|██████████ | 1003/1167 [03:11<00:33, 4.87it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

86%|██████████ | 1004/1167 [03:11<00:32, 4.95it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

86%|██████████ | 1005/1167 [03:12<00:32, 4.92it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

86%|██████████ | 1006/1167 [03:12<00:34, 4.69it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

86%|██████████ | 1007/1167 [03:12<00:33, 4.76it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

86%|██████████ | 1008/1167 [03:12<00:32, 4.88it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

86%|██████████ | 1009/1167 [03:12<00:32, 4.92it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

87%|██████████ | 1010/1167 [03:13<00:31, 4.94it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

87%|██████████ | 1011/1167 [03:13<00:31, 4.94it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

87%|██████████ | 1012/1167 [03:13<00:31, 4.90it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

87%|██████████ | 1013/1167 [03:13<00:32, 4.79it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

87%|██████████ | 1014/1167 [03:14<00:33, 4.59it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

87%|██████████ | 1015/1167 [03:14<00:31, 4.82it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

87%|██████████ | 1016/1167 [03:14<00:30, 4.91it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

87%|██████████ | 1017/1167 [03:14<00:29, 5.07it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

87%|██████████ | 1018/1167 [03:14<00:29, 5.0lit/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

87%|██████████ | 1019/1167 [03:15<00:32, 4.6lit/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

87%|██████████ | 1020/1167 [03:15<00:31, 4.69it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

87%|██████████ | 1021/1167 [03:15<00:31, 4.71it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

88%|██████████ | 1022/1167 [03:15<00:32, 4.46it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

88%|██████████ | 1023/1167 [03:15<00:30, 4.76it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

88%|██████████ | 1024/1167 [03:16<00:30, 4.73it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

88%|██████████ | 1025/1167 [03:16<00:28, 4.92it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

88%|██████████ | 1026/1167 [03:16<00:28, 4.92it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

88%|██████████ | 1027/1167 [03:16<00:27, 5.05it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

88%|██████████ | 1028/1167 [03:16<00:27, 5.12it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

88%|██████████ | 1029/1167 [03:17<00:28, 4.80it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

88%|██████████ | 1030/1167 [03:17<00:28, 4.86it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

88%|██████████| 1031/1167 [03:17<00:27, 4.98it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

88%|██████████ | 1032/1167 [03:17<00:26, 5.03it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

89%|██████████ | 1033/1167 [03:17<00:26, 5.15it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

89%|██████████ | 1034/1167 [03:18<00:26, 4.97it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

89%|██████████ | 1035/1167 [03:18<00:26, 5.02it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

89%|██████████ | 1036/1167 [03:18<00:26, 4.92it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

89%|██████████ | 1037/1167 [03:18<00:27, 4.81it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

89%|██████████ | 1038/1167 [03:18<00:25, 5.01it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

89%|██████████ | 1039/1167 [03:19<00:26, 4.80it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

89%|██████████ | 1040/1167 [03:19<00:25, 5.02it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxFForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

89%|██████████ | 1041/1167 [03:19<00:24, 5.05it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxFForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

89%|██████████ | 1042/1167 [03:19<00:24, 5.01it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

89%|██████████ | 1043/1167 [03:19<00:23, 5.20it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

89%|██████████ | 1044/1167 [03:20<00:25, 4.76it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

90%|██████████ | 1045/1167 [03:20<00:25, 4.75it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

90%|██████████ | 1046/1167 [03:20<00:26, 4.65it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

90%|██████████ | 1047/1167 [03:20<00:24, 4.91it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

90%|██████████ | 1048/1167 [03:20<00:23, 5.06it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

90%|██████████ | 1049/1167 [03:21<00:26, 4.44it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

90%|██████████ | 1050/1167 [03:21<00:24, 4.73it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

90%|██████████ | 1051/1167 [03:21<00:24, 4.68it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

90%|██████████| 1052/1167 [03:21<00:23, 4.84it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

90%|██████████ | 1053/1167 [03:21<00:22, 4.98it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

90%|██████████ | 1054/1167 [03:22<00:22, 4.94it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

90%|██████████| 1055/1167 [03:22<00:22, 5.00it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

90%|██████████ | 1056/1167 [03:22<00:21, 5.22it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

91%|██████████ | 1057/1167 [03:22<00:21, 5.16it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

91%|██████████ | 1058/1167 [03:22<00:22, 4.95it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

91%|██████████ | 1059/1167 [03:23<00:22, 4.84it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

91%|██████████ | 1060/1167 [03:23<00:22, 4.72it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

91%|██████████| 1061/1167 [03:23<00:21, 4.93it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

91%|██████████| 1062/1167 [03:23<00:20, 5.04it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

91%|██████████| 1063/1167 [03:23<00:20, 5.15it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

91%|██████████ | 1064/1167 [03:24<00:20, 4.99it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

91%|██████████ | 1065/1167 [03:24<00:21, 4.79it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

91%|██████████ | 1066/1167 [03:24<00:20, 4.92it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

91%|██████████ | 1067/1167 [03:24<00:20, 4.94it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

92%|██████████ | 1068/1167 [03:25<00:19, 4.96it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

92%|██████████ | 1069/1167 [03:25<00:20, 4.72it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

92%|██████████ | 1070/1167 [03:25<00:20, 4.79it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

92%|██████████| 1071/1167 [03:25<00:19, 4.93it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

92%|██████████ | 1072/1167 [03:25<00:19, 4.81it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

92%|██████████ | 1073/1167 [03:26<00:18, 5.00it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

92%|██████████ | 1074/1167 [03:26<00:18, 4.94it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

92%|██████████ | 1075/1167 [03:26<00:18, 4.98it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

92%|██████████| 1076/1167 [03:26<00:17, 5.08it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

92%|██████████| 1077/1167 [03:26<00:17, 5.14it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

92%|██████████| 1078/1167 [03:27<00:17, 5.15it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

92%|██████████| 1079/1167 [03:27<00:18, 4.85it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

93%|██████████ | 1080/1167 [03:27<00:18, 4.82it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

93%|██████████ | 1081/1167 [03:27<00:18, 4.69it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

93%|██████████ | 1082/1167 [03:27<00:17, 4.81it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

93%|██████████ | 1083/1167 [03:28<00:16, 4.97it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

93%|██████████| 1084/1167 [03:28<00:17, 4.82it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

93%|██████████| 1085/1167 [03:28<00:16, 5.00it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

93%|██████████| 1086/1167 [03:28<00:16, 4.81it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

93%|██████████| 1087/1167 [03:28<00:16, 4.92it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

93%|██████████ | 1088/1167 [03:29<00:15, 4.97it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

93%|██████████ | 1089/1167 [03:29<00:16, 4.85it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

93%|██████████ | 1090/1167 [03:29<00:15, 4.92it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

93%|██████████ | 1091/1167 [03:29<00:14, 5.13it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

94%|██████████| 1092/1167 [03:29<00:14, 5.16it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

94%|██████████| 1093/1167 [03:30<00:15, 4.85it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

94%|██████████| 1094/1167 [03:30<00:15, 4.68it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

94%|██████████| 1095/1167 [03:30<00:14, 4.86it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

94%|██████████ | 1096/1167 [03:30<00:14, 4.95it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

94%|██████████ | 1097/1167 [03:30<00:13, 5.09it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

94%|██████████| 1098/1167 [03:31<00:13, 5.16it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

94%|██████████| 1099/1167 [03:31<00:13, 4.86it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxFForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmForCausalLM', 'Olm2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

94%|██████████| 1100/1167 [03:31<00:14, 4.69it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxFForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmForCausalLM', 'Olm2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

94%|██████████| 1101/1167 [03:31<00:13, 4.95it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

94%|██████████| 1102/1167 [03:31<00:13, 5.00it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

95%|██████████| 1103/1167 [03:32<00:13, 4.87it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

95%|██████████ | 1104/1167 [03:32<00:13, 4.64it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

95%|██████████ | 1105/1167 [03:32<00:12, 4.84it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

95%|██████████| 1106/1167 [03:32<00:12, 5.01it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

95%|██████████| 1107/1167 [03:32<00:12, 4.90it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

95%|██████████| 1108/1167 [03:33<00:12, 4.87it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

95%|██████████| 1109/1167 [03:33<00:11, 4.85it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

95%|██████████| 1110/1167 [03:33<00:11, 4.88it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

95%|██████████| 1111/1167 [03:33<00:11, 4.88it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

95%|██████████| 1112/1167 [03:33<00:11, 4.96it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

95%|██████████| 1113/1167 [03:34<00:10, 4.97it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

95%|██████████| 1114/1167 [03:34<00:11, 4.72it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

96%|██████████| 1115/1167 [03:34<00:10, 4.94it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

96%|██████████| 1116/1167 [03:34<00:10, 4.92it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

96%|██████████| 1117/1167 [03:35<00:10, 4.82it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

96%|██████████| 1118/1167 [03:35<00:10, 4.70it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

96%|██████████| 1119/1167 [03:35<00:10, 4.78it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

96%|██████████| 1120/1167 [03:35<00:09, 4.81it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

96%|██████████| 1121/1167 [03:35<00:09, 4.69it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

96%|██████████| 1122/1167 [03:36<00:09, 4.74it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

96%|██████████| 1123/1167 [03:36<00:08, 4.90it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

96%|██████████| 1124/1167 [03:36<00:08, 4.92it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

96%|██████████| 1125/1167 [03:36<00:08, 4.78it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

96%|██████████| 1126/1167 [03:36<00:08, 4.78it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

97%|██████████| 1127/1167 [03:37<00:08, 4.84it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

97%|██████████| 1128/1167 [03:37<00:08, 4.67it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

97%|██████████| 1129/1167 [03:37<00:08, 4.61it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

97%|██████████| 1130/1167 [03:37<00:07, 4.86it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

97%|██████████| 1131/1167 [03:37<00:07, 4.76it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

97%|██████████| 1132/1167 [03:38<00:07, 4.92it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

97%|██████████| 1133/1167 [03:38<00:07, 4.82it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

97%|██████████| 1134/1167 [03:38<00:06, 4.83it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

97%|██████████| 1135/1167 [03:38<00:06, 4.75it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

97%|██████████| 1136/1167 [03:38<00:06, 4.74it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

97%|██████████| 1137/1167 [03:39<00:06, 4.81it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

98%|██████████| 1138/1167 [03:39<00:05, 4.85it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

98%|██████████| 1139/1167 [03:39<00:05, 4.79it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxFForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmForCausalLM', 'Olm2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

98%|██████████| 1140/1167 [03:39<00:05, 4.83it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxFForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmForCausalLM', 'Olm2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

98%|██████████| 1141/1167 [03:40<00:05, 4.62it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

98%|██████████| 1142/1167 [03:40<00:05, 4.52it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

98%|██████████| 1143/1167 [03:40<00:05, 4.68it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

98%|██████████| 1144/1167 [03:40<00:05, 4.59it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

98%|██████████| 1145/1167 [03:40<00:04, 4.73it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

98%|██████████| 1146/1167 [03:41<00:04, 4.88it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

98%|██████████| 1147/1167 [03:41<00:03, 5.00it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

98%|██████████| 1148/1167 [03:41<00:03, 4.94it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

98%|██████████| 1149/1167 [03:41<00:03, 4.62it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

99%|██████████| 1150/1167 [03:41<00:03, 4.84it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

99%|██████████| 1151/1167 [03:42<00:03, 4.87it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

99%|██████████| 1152/1167 [03:42<00:03, 4.96it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxFForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

99%|██████████| 1153/1167 [03:42<00:02, 4.85it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxFForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

99%|██████████| 1154/1167 [03:42<00:02, 4.85it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

99%|██████████| 1155/1167 [03:42<00:02, 4.81it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

99%|██████████| 1156/1167 [03:43<00:02, 4.55it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

99%|██████████| 1157/1167 [03:43<00:02, 4.45it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

99%|██████████| 1158/1167 [03:43<00:02, 4.39it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

99%|██████████| 1159/1167 [03:43<00:01, 4.41it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

99%|██████████| 1160/1167 [03:44<00:01, 4.45it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

99%|██████████| 1161/1167 [03:44<00:01, 4.55it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'Mus

icgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

100%|██████████| 1162/1167 [03:44<00:01, 4.76it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

100%|██████████| 1163/1167 [03:44<00:00, 4.63it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM',

'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

100%|██████████| 1164/1167 [03:44<00:00, 4.85it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuMForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead

d', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

100%|██████████| 1165/1167 [03:45<00:00, 4.72it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransfoXLMLHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM']].

100%|██████████| 1166/1167 [03:45<00:00, 4.61it/s]Device set to use cuda:0
The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM',

```
'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliumForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoForCausalLM', 'Olmo2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'Starcoder2ForCausalLM', 'TransformoXLLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].
```

```
100%|██████████| 1167/1167 [03:45<00:00, 5.17it/s]
```

Accuracy: 0.955

Accuracy for label Normal: 0.982

Accuracy for label Depression: 0.938

Accuracy for label Anxiety: 0.943

Classification Report:

	precision	recall	f1-score	support
Normal	0.96	0.98	0.97	389
Depression	0.94	0.94	0.94	389
Anxiety	0.97	0.94	0.95	389
micro avg	0.96	0.95	0.96	1167
macro avg	0.96	0.95	0.96	1167
weighted avg	0.96	0.95	0.96	1167

Confusion Matrix:

```
[[382  6  1]
 [ 9 365 12]
 [ 5 17 367]]
```

9. Merge Adaptor with Base Model

```
In [18]: # Cell 2: Load your fine-tuned adapter model
output_dir = "llama-3.1-fine-tuned-model"
model, tokenizer = FastLanguageModel.from_pretrained(
    output_dir,
    device_map="auto",
)

# Cell 3: Merge adapter weights with base model
print("Merging adapter weights with base model...")
```

```
merged_model = model.merge_and_unload()
print("Merge completed!")

# Cell 4: Save the merged model locally
merged_output_dir = "llama-3.1-merged-model"
print(f"Saving merged model to {merged_output_dir}...")
merged_model.save_pretrained(merged_output_dir)
tokenizer.save_pretrained(merged_output_dir)
print("Merged model saved locally!")
```

==(=====)== Unsloth 2025.3.19: Fast Llama patching. Transformers: 4.51.3.
 \ \ /| NVIDIA GeForce RTX 4090. Num GPUs = 1. Max memory: 23.542 GB.
 Platform: Linux.
 0^0/ _/ \ Torch: 2.6.0+cu124. CUDA: 8.9. CUDA Toolkit: 12.4. Triton: 3.
 2.0
 \ _____ / Bfloat16 = TRUE. FA [Xformers = 0.0.29.post3. FA2 = False]
 "-_____" Free license: <http://github.com/unslothai/unsloth>
 Unsloth: Fast downloading is enabled - ignore downloading bars which are red
 colored!
 Merging adapter weights with base model...

```
/home/makers/.pyenv/versions/dion-dsai/lib/python3.12/site-packages/peft/tun
ers/lora/bnb.py:351: UserWarning: Merge lora module to 4-bit linear may get
different generations due to rounding errors.
  warnings.warn(
Merge completed!
Saving merged model to llama-3.1-merged-model...
Merged model saved locally!
```

10. Visualise Fine Tuned Output

```
In [ ]: text = "I'm trapped in a storm of emotions that I can't control, and it feel
prompt = f""Classify the text into Normal, Depression, Anxiety, and return
text: {text}
label: """".strip()

pipe = pipeline(
    "text-generation",
    model=merged_model,
    tokenizer=tokenizer,
    device_map="auto",
)

outputs = pipe(prompt, max_new_tokens=2, do_sample=True, temperature=0.1)
print(outputs[0]["generated_text"].split("label: ")[-1].strip())
```


Device set to use cuda:0

The model 'PeftModelForCausalLM' is not supported for text-generation. Supported models are ['AriaTextForCausalLM', 'BambaForCausalLM', 'BartForCausalLM', 'BertLMHeadModel', 'BertGenerationDecoder', 'BigBirdForCausalLM', 'BigBirdPegasusForCausalLM', 'BioGptForCausalLM', 'BlenderbotForCausalLM', 'BlenderbotSmallForCausalLM', 'BloomForCausalLM', 'CamembertForCausalLM', 'LlamaForCausalLM', 'CodeGenForCausalLM', 'CohereForCausalLM', 'Cohere2ForCausalLM', 'CpmAntForCausalLM', 'CTRLLMHeadModel', 'Data2VecTextForCausalLM', 'DbrxForCausalLM', 'DeepseekV3ForCausalLM', 'DiffLlamaForCausalLM', 'ElectraForCausalLM', 'Emu3ForCausalLM', 'ErnieForCausalLM', 'FalconForCausalLM', 'FalconMambaForCausalLM', 'FuyuForCausalLM', 'GemmaForCausalLM', 'Gemma2ForCausalLM', 'Gemma3ForConditionalGeneration', 'Gemma3ForCausalLM', 'GitForCausalLM', 'GlmForCausalLM', 'Glm4ForCausalLM', 'GotOcr2ForConditionalGeneration', 'GPT2LMHeadModel', 'GPT2LMHeadModel', 'GPTBigCodeForCausalLM', 'GPTNeoForCausalLM', 'GPTNeoXForCausalLM', 'GPTNeoXJapaneseForCausalLM', 'GPTJForCausalLM', 'GraniteForCausalLM', 'GraniteMoeForCausalLM', 'GraniteMoeSharedForCausalLM', 'HeliuForCausalLM', 'JambaForCausalLM', 'JetMoeForCausalLM', 'LlamaForCausalLM', 'Llama4ForCausalLM', 'Llama4ForCausalLM', 'MambaForCausalLM', 'Mamba2ForCausalLM', 'MarianForCausalLM', 'MBartForCausalLM', 'MegaForCausalLM', 'MegatronBertForCausalLM', 'MistralForCausalLM', 'MixtralForCausalLM', 'MllamaForCausalLM', 'MoshiForCausalLM', 'MptForCausalLM', 'MusicgenForCausalLM', 'MusicgenMelodyForCausalLM', 'MvpForCausalLM', 'NemotronForCausalLM', 'OlmoeForCausalLM', 'Olmoe2ForCausalLM', 'OlmoeForCausalLM', 'OpenLlamaForCausalLM', 'OpenAIGPTLMHeadModel', 'OPTForCausalLM', 'PegasusForCausalLM', 'PersimmonForCausalLM', 'PhiForCausalLM', 'Phi3ForCausalLM', 'Phi4MultimodalForCausalLM', 'PhimoeForCausalLM', 'PLBartForCausalLM', 'ProphetNetForCausalLM', 'QDQBertLMHeadModel', 'Qwen2ForCausalLM', 'Qwen2MoeForCausalLM', 'Qwen3ForCausalLM', 'Qwen3MoeForCausalLM', 'RecurrentGemmaForCausalLM', 'ReformerModelWithLMHead', 'RemBertForCausalLM', 'RobertaForCausalLM', 'RobertaPreLayerNormForCausalLM', 'RoCBertForCausalLM', 'RoFormerForCausalLM', 'RwkvForCausalLM', 'Speech2Text2ForCausalLM', 'StableLmForCausalLM', 'StarCoder2ForCausalLM', 'TransformoXLMHeadModel', 'TrOCRForCausalLM', 'WhisperForCausalLM', 'XGLMForCausalLM', 'XLMWithLMHeadModel', 'XLMProphetNetForCausalLM', 'XLMRobertaForCausalLM', 'XLMRobertaXLForCausalLM', 'XLNetLMHeadModel', 'XmodForCausalLM', 'ZambaForCausalLM', 'Zamba2ForCausalLM'].

Depression

11. Push to HuggingFace repo

```
In [ ]: from huggingface_hub import login

# Cell A5: Login to Hugging Face Hub
HF_TOKEN = HF_TOKEN # Replace with your token
login(token=HF_TOKEN)

# Cell 6: Push the merged model to Hugging Face Hub
model_id = "fiendfrye/mental-status-classifier-lama-3.1-8b-fine-tuned" # Re

# Push to hub - this might take a while depending on your connection speed
merged_model.push_to_hub(
    model_id,
    tokenizer=tokenizer,
    private=False, # Set to True if you want a private repository
    commit_message="Upload merged fine-tuned Llama 3.1 8B model. This model"
```

```
)

print(f"Merged model successfully uploaded to https://huggingface.co/{model_

README.md: 0%|          | 0.00/608 [00:00<?, ?B/s]
0%|          | 0/2 [00:00<?, ?it/s]
model-00002-of-00002.safetensors: 0%|          | 0.00/1.05G [00:00<?, ?B/
s]
model-00001-of-00002.safetensors: 0%|          | 0.00/4.65G [00:00<?, ?B/
s]
Saved model to https://huggingface.co/fiendfrye/mental-status-classifier-lam
a-3.1-8b-fine-tuned
Merged model successfully uploaded to https://huggingface.co/fiendfrye/menta
l-status-classifier-lama-3.1-8b-fine-tuned
```

12. Create Model Card

```
In [21]: # Cell 7: Create model card with details about your fine-tuning
model_card = """
---
language:
- en
tags:
- llama
- llama-3.1
- fine-tuned
- text-generation
library_name: transformers
pipeline_tag: text-generation
license: cc-by-nc-4.0
---

# Fine-tuned Llama 3.1 8B

This model is a fine-tuned version of the [Meta Llama 3.1 8B Instruct](https

## Training Details

- **Base Model:** Meta-Llama-3.1-8B-Instruct
- **Fine-tuning Method:** QLoRA with Unsloth optimization
- **Parameters:**
  - LoRA rank: 16
  - LoRA alpha: 16
  - Batch size: 4
  - Learning rate: 2e-4
  - Epochs: 1

## Intended Use

This model is fine-tuned on this [Kaggle Dataset on mental health sentiment
The model classifies the user to 1 of 3 categories; 1. Normal, 2. Depressor

**This was developed as part of a larger project, creating a classification
***This was developed for SC1015 by Chia Dion Yi.
```

Example Usage

```
```python
from transformers import AutoModelForCausalLM, AutoTokenizer
import torch

model_name = "{model_id}"
model = AutoModelForCausalLM.from_pretrained(model_name, torch_dtype=torch.float16)
tokenizer = AutoTokenizer.from_pretrained(model_name)

text = "I'm trapped in a storm of emotions that I can't control, and it feels like I'm drowning."
prompt = f"Classify the text into Normal, Depression, Anxiety, and return the result in JSON format: {{text: {text}, label: \".\".strip()}}\n"

inputs = tokenizer(prompt, return_tensors="pt").to(model.device)

pipe = pipeline(
 "text-generation",
 model=model,
 tokenizer=tokenizer,
 device_map="auto",
)

outputs = pipe(prompt, max_new_tokens=2, do_sample=True, temperature=0.1)
print(outputs[0]["generated_text"].split("label: ")[-1].strip())
```

## Limitations
- Nil

"""

# Save model card
with open("README.md", "w") as f:
    f.write(model_card)

# Push the model card to the hub
from huggingface_hub import HfApi
api = HfApi()
api.upload_file(
    path_or_fileobj="README.md",
    path_in_repo="README.md",
    repo_id=model_id,
    repo_type="model",
)

print("Model card uploaded successfully!")
```

Model card uploaded successfully!