```
!pip install transformers[sentencepiece] datasets sacrebleu rouge_score py7zr -q
```

```
323.25s - pydevd: Sending message related to process being replaced timed-out after 5 seconds
pip: No match.
```

```
!pip install sacrebleu
```

```
huggingface/tokenizers: The current process just got forked, after parallelism has already been used. Disabling parallelism to avoid dea
To disable this warning, you can either:
        - Avoid using `tokenizers` before the fork if possible
        - Explicitly set the environment variable TOKENIZERS_PARALLELISM=(true | false)
1183.11s - pydevd: Sending message related to process being replaced timed-out after 5 seconds
Looking in indexes: https://pypi.org/simple, https://pypi.ngc.nvidia.com
Collecting sacrebleu
  Downloading sacrebleu-2.4.2-py3-none-any.whl.metadata (58 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 58.0/58.0 kB 16.2 MB/s eta 0:00:00
Collecting portalocker (from sacrebleu)
  Downloading portalocker-2.8.2-py3-none-any.whl.metadata (8.5 kB)
Requirement already satisfied: regex in ./envs/default-tensorflow-gpu-2.12.0/lib/python3.10/site-packages (from sacrebleu) (2024.4.16)
Collecting tabulate>=0.8.9 (from sacrebleu)
  Downloading tabulate-0.9.0-py3-none-any.whl.metadata (34 kB)
Requirement already satisfied: numpy>=1.17 in ./envs/default-tensorflow-gpu-2.12.0/lib/python3.10/site-packages (from sacrebleu) (1.26.4
Collecting colorama (from sacrebleu)
  Downloading colorama-0.4.6-py2.py3-none-any.whl.metadata (17 kB)
Requirement already satisfied: lxml in ./envs/default-tensorflow-gpu-2.12.0/lib/python3.10/site-packages (from sacrebleu) (5.2.1)
Downloading sacrebleu-2.4.2-py3-none-any.whl (106 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 106.7/106.7 kB 27.7 MB/s eta 0:00:00
Downloading tabulate-0.9.0-py3-none-any.whl (35 kB)
Downloading colorama-0.4.6-py2.py3-none-any.whl (25 kB)
Downloading portalocker-2.8.2-py3-none-any.whl (17 kB)
Installing collected packages: tabulate, portalocker, colorama, sacrebleu
Successfully installed colorama-0.4.6 portalocker-2.8.2 sacrebleu-2.4.2 tabulate-0.9.0
```

```
!nvidia-smi
```

```
408.08s - pydevd: Sending message related to process being replaced timed-out after 5 seconds
Tue Apr 30 16:27:17 2024
+---------------------------------------------------------------------------------------+
| NVIDIA-SMI 545.23.08              Driver Version: 545.23.08    CUDA Version: 12.3     |
|-----------------------------------------+----------------------+----------------------+
| GPU  Name                 Persistence-M | Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp   Perf          Pwr:Usage/Cap |         Memory-Usage | GPU-Util  Compute M. |
|                                         |                      |               MIG M. |
|=========================================+======================+======================|
|   0  NVIDIA A100 80GB PCIe          On  | 00000000:21:00.0 Off |                    0 |
| N/A   33C    P0              45W / 300W |      4MiB / 81920MiB |      0%      Default |
|                                         |                      |             Disabled |
+-----------------------------------------+----------------------+----------------------+

+---------------------------------------------------------------------------------------+
| Processes:                                                                            |
|  GPU   GI   CI        PID   Type   Process name                            GPU Memory |
|        ID   ID                                                             Usage      |
|=======================================================================================|
|  No running processes found                                                           |
+---------------------------------------------------------------------------------------+
```

Importing all the necessary libraries

```
from transformers import pipeline, set_seed

import matplotlib.pyplot as plt

from datasets import load_dataset, load_metric

import pandas as pd

from transformers import AutoModelForSeq2SeqLM, AutoTokenizer

import nltk

from nltk.tokenize import sent_tokenize

from tqdm import tqdm

import torch
```

```
nltk.download("punkt")
```

```
[nltk_data] Downloading package punkt to /home/mrami010/nltk_data...
[nltk_data]   Package punkt is already up-to-date!
True
```

```
device = "cuda" if torch.cuda.is_available() else "cpu"

model_ckpt = "google/pegasus-cnn_dailymail"

tokenizer = AutoTokenizer.from_pretrained(model_ckpt)

model_pegasus = AutoModelForSeq2SeqLM.from_pretrained(model_ckpt).to(device)
```

```
Some weights of PegasusForConditionalGeneration were not initialized from the model checkpoint at google/pegasus-cnn_dailymail and are n
You should probably TRAIN this model on a down-stream task to be able to use it for predictions and inference.
```

```
device
```

```
'cuda'
```

```
dataset_samsum = load_dataset("samsum")
dataset_samsum
```

```
DatasetDict({
    train: Dataset({
        features: ['id', 'dialogue', 'summary'],
        num_rows: 14732
    })
    test: Dataset({
        features: ['id', 'dialogue', 'summary'],
        num_rows: 819
    })
    validation: Dataset({
        features: ['id', 'dialogue', 'summary'],
        num_rows: 818
    })
})
```

```
split_lengths = [len(dataset_samsum[split]) for split in dataset_samsum]
split_lengths
```

```
[14732, 819, 818]
```

```
print(f"Features: {dataset_samsum['train'].column_names}")
```

```
Features: ['id', 'dialogue', 'summary']
```

```
print("\nDialogue:")

print(dataset_samsum["test"][1]["dialogue"])

print("\nSummary:")

print(dataset_samsum["test"][1]["summary"])
```

```
Dialogue:
Eric: MACHINE!
Rob: That's so gr8!
Eric: I know! And shows how Americans see Russian ;)
Rob: And it's really funny!
Eric: I know! I especially like the train part!
Rob: Hahaha! No one talks to the machine like that!
Eric: Is this his only stand-up?
Rob: Idk. I'll check.
Eric: Sure.
Rob: Turns out no! There are some of his stand-ups on youtube.
Eric: Gr8! I'll watch them now!
Rob: Me too!
Eric: MACHINE!
Rob: MACHINE!
Eric: TTYL?
Rob: Sure :)
```

```
Summary:
Eric and Rob are going to watch a stand-up on youtube.
```

```
dialogue = dataset_samsum["test"][0]["dialogue"]
dialogue
```

```
"Hannah: Hey, do you have Betty's number?\nAmanda: Lemme check\nHannah: <file_gif>\nAmanda: Sorry, can't find it.\nAmanda: Ask
Larry\nAmanda: He called her last time we were at the park together\nHannah: I don't know him well\nHannah: <file_gif>\nAmanda: Don't
be shy, he's very nice\nHannah: If you say so..\nHannah: I'd rather you texted him\nAmanda: Just text him 🙂\nHannah: Urgh..
Alright\nHannah: Bye\nAmanda: Bye bye"
```

```
dialogue_summary = dataset_samsum["test"][0]["summary"]
dialogue_summary
```

```
"Hannah needs Betty's number but Amanda doesn't have it. She needs to contact Larry."
```

```
pipe = pipeline('summarization' , model = model_ckpt)
```

```
Some weights of PegasusForConditionalGeneration were not initialized from the model checkpoint at google/pegasus-cnn_dailymail and are n
You should probably TRAIN this model on a down-stream task to be able to use it for predictions and inference.
```

```
pipe_out = pipe(dialogue)
pipe_out
```

```
Your max_length is set to 128, but your input_length is only 122. Since this is a summarization task, where outputs shorter than the inp
[{'summary_text': "Amanda: Ask Larry Amanda: He called her last time we were at the park together .<n>Hannah: I'd rather you texted him
.<n>Amanda: Just text him ."}]
```

```
print(pipe_out[0]['summary_text'].replace(" .<n>", ".\n"))
```

```
Amanda: Ask Larry Amanda: He called her last time we were at the park together.
Hannah: I'd rather you texted him.
Amanda: Just text him .
```

```
def generate_batch_sized_chunks(list_of_elements, batch_size):
    """split the dataset into smaller batches that we can process simultaneously
    Yield successive batch-sized chunks from list_of_elements."""
    for i in range(0, len(list_of_elements), batch_size):
        yield list_of_elements[i : i + batch_size]
```

```
def calculate_metric_on_test_ds(dataset, metric, model, tokenizer,
                                batch_size=16, device=device,
                                column_text="article",
                                column_summary="highlights"):
    article_batches = list(generate_batch_sized_chunks(dataset[column_text], batch_size))
    target_batches = list(generate_batch_sized_chunks(dataset[column_summary], batch_size))

    for article_batch, target_batch in tqdm(
        zip(article_batches, target_batches), total=len(article_batches)):

        inputs = tokenizer(article_batch, max_length=1024,  truncation=True,
                        padding="max_length", return_tensors="pt")

        summaries = model.generate(input_ids=inputs["input_ids"].to(device),
                         attention_mask=inputs["attention_mask"].to(device),
                         length_penalty=0.8, num_beams=8, max_length=128)
        ''' parameter for length penalty ensures that the model does not generate sequences that are too long. '''

        decoded_summaries = [tokenizer.decode(s, skip_special_tokens=True,
                                clean_up_tokenization_spaces=True)
               for s in summaries]

        decoded_summaries = [d.replace("", " ") for d in decoded_summaries]


        metric.add_batch(predictions=decoded_summaries, references=target_batch)

    score = metric.compute()
    return score
```

```
rouge_metric = load_metric('rouge')

score = calculate_metric_on_test_ds(dataset_samsum['test'], rouge_metric, model_pegasus, tokenizer, column_text = 'dialogue', column_summary
```

```
/tmp/ipykernel_2152727/1560040859.py:1: FutureWarning: load_metric is deprecated and wil
  rouge_metric = load_metric('rouge')
/home/mrami010/envs/default-tensorflow-gpu-2.12.0/lib/python3.10/site-packages/datasets/
You can avoid this message in future by passing the argument `trust_remote_code=True`.
Passing `trust_remote_code=True` will be mandatory to load this metric from the next maj
  warnings.warn(
Downloading builder script:   0%|          | 0.00/2.17k [00:00<?, ?B/s]
100%|██████████| 103/103 [05:40<00:00,  3.30s/it]
```

```
score
```

```
{'rouge1': AggregateScore(low=Score(precision=0.008594284774101471, recall=0.05485776073298719, fmeasure=0.01447275460524098),
mid=Score(precision=0.009252590835506039, recall=0.05919817712317914, fmeasure=0.015555281395557525),
high=Score(precision=0.009912824487290148, recall=0.06298211061176136, fmeasure=0.016616307151489346)),
 'rouge2': AggregateScore(low=Score(precision=8.328203879245665e-05, recall=0.0005408533120551895, fmeasure=0.00014159952790986042),
mid=Score(precision=0.00017567194816082564, recall=0.0011518683051349819, fmeasure=0.00029762747810706334),
high=Score(precision=0.00028518215847744 3, recall=0.0019022781597030538, fmeasure=0.000477802024515758)),
 'rougeL': AggregateScore(low=Score(precision=0.008581775890773, recall=0.05532231861412368, fmeasure=0.014450546308416329),
mid=Score(precision=0.009226558098950764, recall=0.058918321473448235, fmeasure=0.015504705210147636),
high=Score(precision=0.00991667963696394, recall=0.06281326104073413, fmeasure=0.016626901674327812)),
 'rougeLsum': AggregateScore(low=Score(precision=0.008624203077807606, recall=0.05539753598135289, fmeasure=0.014550923923851633),
mid=Score(precision=0.009271219797657888, recall=0.05902633137607505, fmeasure=0.015588996153605775),
high=Score(precision=0.009933891323429105, recall=0.06262557297464545, fmeasure=0.016675093086274868))}
```

```
import torch
from transformers import PegasusForConditionalGeneration, PegasusTokenizer
import sacrebleu
hypotheses = [item['summary_text'] for item in pipe_out]
bleu_score = sacrebleu.raw_corpus_bleu(hypotheses, dialogue_summary, .01).score
```

```
rouge_names = ["rouge1", "rouge2", "rougeL", "rougeLsum"]
rouge_dict = dict((rn, score[rn].mid.fmeasure ) for rn in rouge_names )

pd.DataFrame(rouge_dict, index = ['pegasus'])
```

	rouge1	rouge2	rougeL	rougeLsum
pegasus	0.015555	0.000298	0.015505	0.015589

```
def convert_examples_to_features(example_batch):
    input_encodings = tokenizer(example_batch['dialogue'] , max_length = 1024, truncation = True )

    with tokenizer.as_target_tokenizer():
        target_encodings = tokenizer(example_batch['summary'], max_length = 128, truncation = True )

    return {
        'input_ids' : input_encodings['input_ids'],
        'attention_mask': input_encodings['attention_mask'],
        'labels': target_encodings['input_ids']
    }
dataset_samsum_pt = dataset_samsum.map(convert_examples_to_features, batched = True)
```

```
Map:   0%|          | 0/14732 [00:00<?, ? examples/s]
/home/mrami010/envs/default-tensorflow-gpu-2.12.0/lib/python3.10/site-packages/transform
  warnings.warn(
/home/mrami010/envs/default-tensorflow-gpu-2.12.0/lib/python3.10/site-packages/transform
  warnings.warn(
/home/mrami010/envs/default-tensorflow-gpu-2.12.0/lib/python3.10/site-packages/transform
```

```
dataset_samsum_pt["train"][0]
```

```
{'id': '13818513',
 'dialogue': "Amanda: I baked  cookies. Do you want some?\r\nJerry: Sure!\r\nAmanda: I'll bring you tomorrow :-)",
 'summary': 'Amanda baked cookies and will bring Jerry some tomorrow.',
 'input_ids': [12195,
  151,
  125,
```

```
 7091,
 3659,
 107,
 842,
 119,
 245,
 181,
 152,
 10508,
 151,
 7435,
 147,
 12195,
 151,
 125,
 131,
 267,
 650,
 119,
 3469,
 29344,
 1],
'attention_mask': [1,
 1,
 1,
 1,
 1,
 1,
 1,
 1,
 1,
 1,
 1,
 1,
 1,
 1,
 1,
 1,
 1,
 1,
 1,
 1,
 1,
 1,
 1,
 1,
 1],
'labels': [12195, 7091, 3659, 111, 138, 650, 10508, 181, 3469, 107, 1]}
```

```
from transformers import DataCollatorForSeq2Seq

seq2seq_data_collator = DataCollatorForSeq2Seq(tokenizer, model=model_pegasus)
```

1. output_dir: This argument specifies the directory where model checkpoints and other outputs will be saved during training.
2. num_train_epochs: It determines the number of times the entire training dataset will be passed forward and backward through the model during training.
3. warmup_steps: This specifies the number of steps during which the learning rate will be increased linearly from 0 to its maximum value. It helps stabilize training and prevent divergence.
4. per_device_train_batch_size: This sets the batch size (number of training examples) processed on each device (like GPU or TPU) during training. per_device_eval_batch_size: Similar to per_device_train_batch_size, but for evaluation data.
5. weight_decay: This is a regularization parameter that penalizes large weights in the model to prevent overfitting.
6. logging_steps: Determines the frequency (in steps) at which logs (like loss values) will be printed to the console during training.
7. evaluation_strategy: Specifies when evaluation will be performed during training. Here, it's set to 'steps', meaning evaluation will be performed at regular intervals specified by eval_steps.
8. eval_steps: The interval (in steps) at which evaluation will be performed during training.
9. save_steps: Determines the frequency (in steps) at which model checkpoints will be saved during training. Setting it to a very large value effectively disables intermediate saving.
10. gradient_accumulation_steps: This parameter specifies the number of gradient

```
!pip install transformers[torch]
```

```
huggingface/tokenizers: The current process just got forked, after parallelism has already been used. Disabling parallelism to avoid dea
To disable this warning, you can either:
        - Avoid using `tokenizers` before the fork if possible
        - Explicitly set the environment variable TOKENIZERS_PARALLELISM=(true | false)
1451.93s - pydevd: Sending message related to process being replaced timed-out after 5 seconds
pip: No match.
```

```
!pip install accelerate -U
```

```
huggingface/tokenizers: The current process just got forked, after parallelism has already been used. Disabling parallelism to avoid dea
To disable this warning, you can either:
        - Avoid using `tokenizers` before the fork if possible
        - Explicitly set the environment variable TOKENIZERS_PARALLELISM=(true | false)
1461.83s - pydevd: Sending message related to process being replaced timed-out after 5 seconds
Looking in indexes: https://pypi.org/simple, https://pypi.ngc.nvidia.com
Requirement already satisfied: accelerate in ./envs/default-tensorflow-gpu-2.12.0/lib/python3.10/site-packages (0.29.3)
Requirement already satisfied: numpy>=1.17 in ./envs/default-tensorflow-gpu-2.12.0/lib/python3.10/site-packages (from accelerate) (1.26.
Requirement already satisfied: packaging>=20.0 in ./envs/default-tensorflow-gpu-2.12.0/lib/python3.10/site-packages (from accelerate) (2
Requirement already satisfied: psutil in ./envs/default-tensorflow-gpu-2.12.0/lib/python3.10/site-packages (from accelerate) (5.9.8)
Requirement already satisfied: pyyaml in ./envs/default-tensorflow-gpu-2.12.0/lib/python3.10/site-packages (from accelerate) (6.0.1)
Requirement already satisfied: torch>=1.10.0 in ./envs/default-tensorflow-gpu-2.12.0/lib/python3.10/site-packages (from accelerate) (2.3
Requirement already satisfied: huggingface-hub in ./envs/default-tensorflow-gpu-2.12.0/lib/python3.10/site-packages (from accelerate) (0
Requirement already satisfied: safetensors>=0.3.1 in ./envs/default-tensorflow-gpu-2.12.0/lib/python3.10/site-packages (from accelerate)
Requirement already satisfied: filelock in ./envs/default-tensorflow-gpu-2.12.0/lib/python3.10/site-packages (from torch>=1.10.0->accele
Requirement already satisfied: typing-extensions>=4.8.0 in ./envs/default-tensorflow-gpu-2.12.0/lib/python3.10/site-packages (from torch
Requirement already satisfied: sympy in ./envs/default-tensorflow-gpu-2.12.0/lib/python3.10/site-packages (from torch>=1.10.0->accelerat
Requirement already satisfied: networkx in ./envs/default-tensorflow-gpu-2.12.0/lib/python3.10/site-packages (from torch>=1.10.0->accele
Requirement already satisfied: jinja2 in ./envs/default-tensorflow-gpu-2.12.0/lib/python3.10/site-packages (from torch>=1.10.0->accelera
Requirement already satisfied: fsspec in ./envs/default-tensorflow-gpu-2.12.0/lib/python3.10/site-packages (from torch>=1.10.0->accelera
Requirement already satisfied: nvidia-cuda-nvrtc-cu12==12.1.105 in ./envs/default-tensorflow-gpu-2.12.0/lib/python3.10/site-packages (fr
Requirement already satisfied: nvidia-cuda-runtime-cu12==12.1.105 in ./envs/default-tensorflow-gpu-2.12.0/lib/python3.10/site-packages (
Requirement already satisfied: nvidia-cuda-cupti-cu12==12.1.105 in ./envs/default-tensorflow-gpu-2.12.0/lib/python3.10/site-packages (fr
Requirement already satisfied: nvidia-cudnn-cu12==8.9.2.26 in ./envs/default-tensorflow-gpu-2.12.0/lib/python3.10/site-packages (from tc
Requirement already satisfied: nvidia-cublas-cu12==12.1.3.1 in ./envs/default-tensorflow-gpu-2.12.0/lib/python3.10/site-packages (from t
Requirement already satisfied: nvidia-cufft-cu12==11.0.2.54 in ./envs/default-tensorflow-gpu-2.12.0/lib/python3.10/site-packages (from t
Requirement already satisfied: nvidia-curand-cu12==10.3.2.106 in ./envs/default-tensorflow-gpu-2.12.0/lib/python3.10/site-packages (from
Requirement already satisfied: nvidia-cusolver-cu12==11.4.5.107 in ./envs/default-tensorflow-gpu-2.12.0/lib/python3.10/site-packages (fr
Requirement already satisfied: nvidia-cusparse-cu12==12.1.0.106 in ./envs/default-tensorflow-gpu-2.12.0/lib/python3.10/site-packages (fr
Requirement already satisfied: nvidia-nccl-cu12==2.20.5 in ./envs/default-tensorflow-gpu-2.12.0/lib/python3.10/site-packages (from torch
Requirement already satisfied: nvidia-nvtx-cu12==12.1.105 in ./envs/default-tensorflow-gpu-2.12.0/lib/python3.10/site-packages (from tor
Requirement already satisfied: triton==2.3.0 in ./envs/default-tensorflow-gpu-2.12.0/lib/python3.10/site-packages (from torch>=1.10.0->a
Requirement already satisfied: nvidia-nvjitlink-cu12 in ./envs/default-tensorflow-gpu-2.12.0/lib/python3.10/site-packages (from nvidia-c
Requirement already satisfied: requests in ./envs/default-tensorflow-gpu-2.12.0/lib/python3.10/site-packages (from huggingface-hub->acce
Requirement already satisfied: tqdm>=4.42.1 in ./envs/default-tensorflow-gpu-2.12.0/lib/python3.10/site-packages (from huggingface-hub->
Requirement already satisfied: MarkupSafe>=2.0 in ./envs/default-tensorflow-gpu-2.12.0/lib/python3.10/site-packages (from jinja2->torch>
Requirement already satisfied: charset-normalizer<4,>=2 in ./envs/default-tensorflow-gpu-2.12.0/lib/python3.10/site-packages (from reque
Requirement already satisfied: idna<4,>=2.5 in ./envs/default-tensorflow-gpu-2.12.0/lib/python3.10/site-packages (from requests->hugging
Requirement already satisfied: urllib3<3,>=1.21.1 in ./envs/default-tensorflow-gpu-2.12.0/lib/python3.10/site-packages (from requests->h
Requirement already satisfied: certifi>=2017.4.17 in ./envs/default-tensorflow-gpu-2.12.0/lib/python3.10/site-packages (from requests->h
Requirement already satisfied: mpmath>=0.19 in ./envs/default-tensorflow-gpu-2.12.0/lib/python3.10/site-packages (from sympy->torch>=1.1
```

```
#!pip uninstall transformers
#!pip uninstall accelerate
#!pip install transformers[torch]
```

```
from transformers import TrainingArguments, Trainer

trainer_args = TrainingArguments(
    output_dir='pegasus-samsum', num_train_epochs=1, warmup_steps=500,
    per_device_train_batch_size=1, per_device_eval_batch_size=1,
    weight_decay=0.01, logging_steps=10,
    evaluation_strategy='steps', eval_steps=500, save_steps=1e6,
    gradient_accumulation_steps=16
)
```

```
trainer = Trainer(model=model_pegasus, args=trainer_args,
                  tokenizer=tokenizer, data_collator=seq2seq_data_collator,
                  train_dataset=dataset_samsum_pt["train"],
                  eval_dataset=dataset_samsum_pt["validation"])
```

```
/home/mrami010/envs/default-tensorflow-gpu-2.12.0/lib/python3.10/site-packages/accelerate/accelerator.py:436: FutureWarning: Passing the
dataloader_config = DataLoaderConfiguration(dispatch_batches=None, split_batches=False)
  warnings.warn(
WARNING:accelerate.utils.other:Detected kernel version 4.18.0, which is below the recommended minimum of 5.5.0; this can cause the proce
```

```
trainer.train()
```

[920/920 42:12, Epoch 0/1]

Step	Training Loss	Validation Loss
500	1.660900	1.484190

```
TrainOutput(global_step=920, training_loss=1.8348752477894659, metrics={'train_runtime': 2535.7401, 'train_samples_per_second': 5.81,
'train_steps_per_second': 0.363, 'total_flos': 5526698901602304.0, 'train_loss': 1.8348752477894659, 'epoch': 1.0})
```

```
# Saving the model

model_pegasus.save_pretrained("pegasus-samsum-model")
```

```
Some non-default generation parameters are set in the model config. These should go into a GenerationConfig file (https://huggingface.co
Non-default generation parameters: {'max_length': 128, 'min_length': 32, 'num_beams': 8, 'length_penalty': 0.8, 'forced_eos_token_id': 1
```

```
tokenizer.save_pretrained("tokenizer")
```

```
('tokenizer/tokenizer_config.json',
 'tokenizer/special_tokens_map.json',
 'tokenizer/spiece.model',
 'tokenizer/added_tokens.json',
 'tokenizer/tokenizer.json')
```

```
dataset_samsum = load_dataset("samsum")
```

```
tokenizer = AutoTokenizer.from_pretrained("tokenizer")
```

```
sample_text = dataset_samsum["test"][0]["dialogue"]

reference = dataset_samsum["test"][0]["summary"]
```

```
gen_kwargs = {"length_penalty": 0.8, "num_beams":8, "max_length": 128}

pipe = pipeline("summarization", model="pegasus-samsum-model",tokenizer=tokenizer)
```

```
print("Dialogue:")
print(sample_text)


print("\nReference Summary:")
print(reference)


print("\nModel Summary:")
print(pipe(sample_text, **gen_kwargs)[0]["summary_text"])
```

```
Your max_length is set to 128, but your input_length is only 122. Since this is a summarization task, where outputs shorter than the inp
Dialogue:
Hannah: Hey, do you have Betty's number?
Amanda: Lemme check
Hannah: <file_gif>
Amanda: Sorry, can't find it.
Amanda: Ask Larry
Amanda: He called her last time we were at the park together
Hannah: I don't know him well
Hannah: <file_gif>
Amanda: Don't be shy, he's very nice
Hannah: If you say so..
Hannah: I'd rather you texted him
Amanda: Just text him 😊
Hannah: Urgh.. Alright
Hannah: Bye
Amanda: Bye bye

Reference Summary:
Hannah needs Betty's number but Amanda doesn't have it. She needs to contact Larry.
```

```
Model Summary:
Amanda can't find Betty's number. Larry called Betty last time they were at the park together. Hannah wants Amanda to text Larry. Amanda
```

```
def generate_batch_sized_chunks(list_of_elements, batch_size):
    """split the dataset into smaller batches that we can process simultaneously
    Yield successive batch-sized chunks from list_of_elements."""
    for i in range(0, len(list_of_elements), batch_size):
        yield list_of_elements[i : i + batch_size]
```

```
def calculate_metric_on_test_ds(dataset, metric, model, tokenizer,
                                batch_size=16, device=device,
                                column_text="article",
                                column_summary="highlights"):
    article_batches = list(generate_batch_sized_chunks(dataset[column_text], batch_size))
    target_batches = list(generate_batch_sized_chunks(dataset[column_summary], batch_size))

    for article_batch, target_batch in tqdm(
        zip(article_batches, target_batches), total=len(article_batches)):

        inputs = tokenizer(article_batch, max_length=1024,  truncation=True,
                        padding="max_length", return_tensors="pt")

        summaries = model.generate(input_ids=inputs["input_ids"].to(device),
                         attention_mask=inputs["attention_mask"].to(device),
                         length_penalty=0.8, num_beams=8, max_length=128)
        ''' parameter for length penalty ensures that the model does not generate sequences that are too long. '''

        decoded_summaries = [tokenizer.decode(s, skip_special_tokens=True,
                                clean_up_tokenization_spaces=True)
               for s in summaries]

        decoded_summaries = [d.replace("", " ") for d in decoded_summaries]


        metric.add_batch(predictions=decoded_summaries, references=target_batch)

    score = metric.compute()
    return score
```

```
torch.cuda.empty_cache()
```

```
rouge_metric = load_metric('rouge')
score = calculate_metric_on_test_ds(dataset_samsum['test'], rouge_metric, model_pegasus, tokenizer, column_text = 'dialogue', column_summary
```

```
/home/mrami010/envs/default-tensorflow-gpu-2.12.0/lib/python3.10/site-packages/datasets/load.py:753: FutureWarning: The repository for r
You can avoid this message in future by passing the argument `trust_remote_code=True`.
Passing `trust_remote_code=True` will be mandatory to load this metric from the next major release of `datasets`.
  warnings.warn(
100%|██████████| 103/103 [03:33<00:00,  2.07s/it]
```

```
print(score)
```

```
{'rouge1': AggregateScore(low=Score(precision=0.010499666227875395, recall=0.0567295507063052, fmeasure=0.01716914284237491), mid=Score(
```

```
rouge_names = ["rouge1", "rouge2", "rougeL", "rougeLsum"]
rouge_dict = dict((rn, score[rn].mid.fmeasure ) for rn in rouge_names )

pd.DataFrame(rouge_dict, index = ['pegasus'])
```

	rouge1	rouge2	rougeL	rougeLsum
pegasus	0.018453	0.000287	0.018316	0.018351

```
import torch
from transformers import PegasusForConditionalGeneration, PegasusTokenizer
import sacrebleu
bleu_score = sacrebleu.raw_corpus_bleu([pipe(sample_text, **gen_kwargs)[0]["summary_text"]], [[reference]], .01).score
```

```
Your max_length is set to 128, but your input_length is only 122. Since this is a summarization task, where outputs shorter than the inp
```

```
print(bleu_score)
```

```
0.1932099074568177
```

```
import os
import sys
from tempfile import NamedTemporaryFile
from urllib.request import urlopen
from urllib.parse import unquote, urlparse
from urllib.error import HTTPError
from zipfile import ZipFile
import tarfile
import shutil

CHUNK_SIZE = 40960
DATA_SOURCE_MAPPING = 'samsum-dataset-text-summarization:https%3A%2F%2Fstorage.googleapis.com%2Fkaggle-data-sets%2F3438844%2F6004344%2Fbundle%

KAGGLE_INPUT_PATH='/kaggle/input'
KAGGLE_WORKING_PATH='/kaggle/working'
KAGGLE_SYMLINK='kaggle'

!umount /kaggle/input/ 2> /dev/null
shutil.rmtree('/kaggle/input', ignore_errors=True)
os.makedirs(KAGGLE_INPUT_PATH, 0o777, exist_ok=True)
os.makedirs(KAGGLE_WORKING_PATH, 0o777, exist_ok=True)

try:
  os.symlink(KAGGLE_INPUT_PATH, os.path.join("..", 'input'), target_is_directory=True)
except FileExistsError:
  pass
try:
  os.symlink(KAGGLE_WORKING_PATH, os.path.join("..", 'working'), target_is_directory=True)
except FileExistsError:
  pass

for data_source_mapping in DATA_SOURCE_MAPPING.split(','):
    directory, download_url_encoded = data_source_mapping.split(':')
    download_url = unquote(download_url_encoded)
    filename = urlparse(download_url).path
    destination_path = os.path.join(KAGGLE_INPUT_PATH, directory)
    try:
        with urlopen(download_url) as fileres, NamedTemporaryFile() as tfile:
            total_length = fileres.headers['content-length']
            print(f'Downloading {directory}, {total_length} bytes compressed')
            dl = 0
            data = fileres.read(CHUNK_SIZE)
            while len(data) > 0:
                dl += len(data)
                tfile.write(data)
                done = int(50 * dl / int(total_length))
                sys.stdout.write(f"\r[{'=' * done}{' ' * (50-done)}] {dl} bytes downloaded")
                sys.stdout.flush()
                data = fileres.read(CHUNK_SIZE)
            if filename.endswith('.zip'):
              with ZipFile(tfile) as zfile:
                zfile.extractall(destination_path)
            else:
              with tarfile.open(tfile.name) as tarfile:
                tarfile.extractall(destination_path)
            print(f'\nDownloaded and uncompressed: {directory}')
    except HTTPError as e:
        print(f'Failed to load (likely expired) {download_url} to path {destination_path}')
        continue
    except OSError as e:
        print(f'Failed to load {download_url} to path {destination_path}')
        continue

print('Data source import complete.')
```

```
Downloading samsum-dataset-text-summarization, 8377572 bytes compressed
[==================================================] 8377572 bytes downloaded
Downloaded and uncompressed: samsum-dataset-text-summarization
Data source import complete.
```

```
!nvidia-smi
```

```
Wed May  1 01:48:06 2024
+---------------------------------------------------------------------------------------+
| NVIDIA-SMI 535.104.05             Driver Version: 535.104.05   CUDA Version: 12.2     |
```

```
|-----------------------------------------+----------------------+----------------------+
| GPU  Name                 Persistence-M | Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp   Perf          Pwr:Usage/Cap |         Memory-Usage | GPU-Util  Compute M. |
|                                         |                      |               MIG M. |
|=========================================+======================+======================|
|   0  Tesla T4                       Off | 00000000:00:04.0 Off |                    0 |
| N/A   50C    P8              10W /  70W |      0MiB / 15360MiB |      0%      Default |
|                                         |                      |                  N/A |
+-----------------------------------------+----------------------+----------------------+

+---------------------------------------------------------------------------------------+
| Processes:                                                                            |
|  GPU   GI   CI        PID   Type   Process name                            GPU Memory |
|        ID   ID                                                             Usage      |
|=======================================================================================|
|  No running processes found                                                           |
+---------------------------------------+-----------------------------------------------+
```

```
%%capture --no-stderr
!pip install transformers==4.37.2
!pip install datasets==2.17.0
!pip install evaluate==0.4.1
!pip install rouge-score==0.1.2
```

```
!transformers-cli env
```

```
2024-05-01 01:48:56.650774: E external/local_xla/xla/stream_executor/cuda/cuda_dnn.cc:9261] Unable to register cuDNN factory: Attempting
2024-05-01 01:48:56.650840: E external/local_xla/xla/stream_executor/cuda/cuda_fft.cc:607] Unable to register cuFFT factory: Attempting
2024-05-01 01:48:56.652808: E external/local_xla/xla/stream_executor/cuda/cuda_blas.cc:1515] Unable to register cuBLAS factory: Attempti
2024-05-01 01:48:59.820473: W tensorflow/compiler/tf2tensorrt/utils/py_utils.cc:38] TF-TRT Warning: Could not find TensorRT
WARNING:tensorflow:From /usr/local/lib/python3.10/dist-packages/transformers/commands/env.py:100: is_gpu_available (from tensorflow.pyth
Instructions for updating:
Use `tf.config.list_physical_devices('GPU')` instead.
2024-05-01 01:49:11.007130: W tensorflow/core/common_runtime/gpu/gpu_bfc_allocator.cc:47] Overriding orig_value setting because the TF_F

Copy-and-paste the text below in your GitHub issue and FILL OUT the two last points.

- `transformers` version: 4.37.2
- Platform: Linux-6.1.58+-x86_64-with-glibc2.35
- Python version: 3.10.12
- Huggingface_hub version: 0.20.3
- Safetensors version: 0.4.3
- Accelerate version: 0.21.0
- Accelerate config:    not found
- PyTorch version (GPU?): 2.2.1+cu121 (True)
- Tensorflow version (GPU?): 2.15.0 (True)
- Flax version (CPU?/GPU?/TPU?): 0.8.2 (gpu)
- Jax version: 0.4.26
- JaxLib version: 0.4.26
- Using GPU in script?: <fill in>
- Using distributed or parallel set-up in script?: <fill in>
```

```
import os
from huggingface_hub import login

login(token="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")

os.environ['MODEL']='facebook/bart-large-xsum'

os.environ["WANDB_API_KEY"]="6d2687ddec0c7690b65db8c705bbcb88ed85dbd1"
os.environ["WANDB_PROJECT"] = "Fine-tuning BART Series LLMs"
os.environ["WANDB_NOTES"] = ""
os.environ["WANDB_NAME"] = "ft-facebook-bart-large-xsum-on-samsum"
```

```
Token will not been saved to git credential helper. Pass `add_to_git_credential=True` if you want to set the git credential as well.
Token is valid (permission: write).
Your token has been saved to /root/.cache/huggingface/token
Login successful
```

```
import warnings

warnings.filterwarnings('ignore')
```

```
import pandas as pd

train=pd.read_csv('/kaggle/input/samsum-dataset-text-summarization/samsum-train.csv')
test=pd.read_csv('/kaggle/input/samsum-dataset-text-summarization/samsum-test.csv')
val=pd.read_csv('/kaggle/input/samsum-dataset-text-summarization/samsum-validation.csv')
type(train)
```

```
pandas.core.frame.DataFrame
def __init__(data=None, index: Axes | None=None, columns: Axes | None=None, dtype:
Dtype | None=None, copy: bool | None=None) -> None

Two-dimensional, size-mutable, potentially heterogeneous tabular data.

Data structure also contains labeled axes (rows and columns).
Arithmetic operations align on both row and column labels. Can be
thought of as a dict-like container for Series objects. The primary
```

```
print(train['dialogue'].iloc[14727])
```

```
Romeo: You are on my ‘People you may know’ list.
Greta: Ah, maybe it is because of the changed number of somebody’s?
Greta: I don’t know you?
Romeo: This might be the beginning of a beautiful relationship
Romeo: How about adding me on your friend list and talk a bit?
Greta: No.
Romeo: Okay I see.
```

```
import re

def clean_tags(text):
    clean=re.compile('<.*?>')
    clean=re.sub(clean, '', text)

    clean='\n'.join([line for line in clean.split('\n') if not re.match('.*:\s*$', line)])
    return clean

test1=clean_tags(train['dialogue'].iloc[14727])
test2=clean_tags(test['dialogue'].iloc[0])

print(test1)
print('\n'*3)
print(test2)
```

```
Romeo: You are on my ‘People you may know’ list.
Greta: Ah, maybe it is because of the changed number of somebody’s?
Greta: I don’t know you?
Romeo: This might be the beginning of a beautiful relationship
Romeo: How about adding me on your friend list and talk a bit?
Greta: No.
Romeo: Okay I see.
```

```
Hannah: Hey, do you have Betty's number?
Amanda: Lemme check
Amanda: Sorry, can't find it.
Amanda: Ask Larry
Amanda: He called her last time we were at the park together
Hannah: I don't know him well
Amanda: Don't be shy, he's very nice
Hannah: If you say so..
Hannah: I'd rather you texted him
Amanda: Just text him 🙄
Hannah: Urgh.. Alright
Hannah: Bye
Amanda: Bye bye
```

```
def clean_df(df, cols):
    for col in cols:
        df[col]=df[col].fillna('').apply(clean_tags)
    return df

train=clean_df(train, ['dialogue','summary'])
test=clean_df(test, ['dialogue', 'summary'])
val=clean_df(val, ['dialogue', 'summary'])

# visualizing results
train.tail(3)
```

	id	dialogue	summary
14729	13819050	John: Every day some bad news. Japan will hunt...	Japan is going to hunt whales again. Island an...
14730	13828395	Jennifer: Dear Celia! How are you doing? \r\nJe...	Celia couldn't make it to the afternoon with t...

```
from datasets import Dataset

train_ds=Dataset.from_pandas(train)
test_ds=Dataset.from_pandas(test)
val_ds=Dataset.from_pandas(val)

print(train_ds)
print('\n'*2)
print(test_ds)
print('\n'*2)
print(val_ds)
```

```
Dataset({
    features: ['id', 'dialogue', 'summary'],
    num_rows: 14732
})



Dataset({
    features: ['id', 'dialogue', 'summary'],
    num_rows: 819
})



Dataset({
    features: ['id', 'dialogue', 'summary'],
    num_rows: 818
})
```

```
train_ds[0]
```

```
{'id': '13818513',
 'dialogue': "Amanda: I baked  cookies. Do you want some?\r\nJerry: Sure!\r\nAmanda: I'll bring you tomorrow :-)",
 'summary': 'Amanda baked cookies and will bring Jerry some tomorrow.'}
```

```
from transformers import BartTokenizer, BartForConditionalGeneration # BERT Tokenizer and architecture

tokenizer=BartTokenizer.from_pretrained(os.getenv('MODEL'))
tokenizer
```

```
BartTokenizer(name_or_path='facebook/bart-large-xsum', vocab_size=50265, model_max_length=1024, is_fast=False, padding_side='right',
truncation_side='right', special_tokens={'bos_token': '<s>', 'eos_token': '</s>', 'unk_token': '<unk>', 'sep_token': '</s>',
'pad_token': '<pad>', 'cls_token': '<s>', 'mask_token': '<mask>'}, clean_up_tokenization_spaces=True),  added_tokens_decoder={
        0: AddedToken("<s>", rstrip=False, lstrip=False, single_word=False, normalized=True, special=True),
        1: AddedToken("<pad>", rstrip=False, lstrip=False, single_word=False, normalized=True, special=True),
        2: AddedToken("</s>", rstrip=False, lstrip=False, single_word=False, normalized=True, special=True),
        3: AddedToken("<unk>", rstrip=False, lstrip=False, single_word=False, normalized=True, special=True),
        50264: AddedToken("<mask>", rstrip=False, lstrip=True, single_word=False, normalized=True, special=True),
}
```

```
def preprocess_func(example):
    inputs=[doc for doc in example['dialogue']]
    model_inputs=tokenizer(inputs, max_length=1024, truncation=True)

    with tokenizer.as_target_tokenizer():
        labels=tokenizer(example['summary'], max_length=128, truncation=True)

    model_inputs['labels']=labels['input_ids']
    return model_inputs

tokenized_train= train_ds.map(preprocess_func, batched=True, remove_columns=['id', 'dialogue', 'summary'])
tokenized_test=test_ds.map(preprocess_func, batched=True, remove_columns=['id', 'dialogue', 'summary'])
tokenized_val=val_ds.map(preprocess_func, batched=True, remove_columns=['id', 'dialogue', 'summary'])

print(tokenized_train)
print(tokenized_test)
print(tokenized_val)
```

```
Map: 100%  14732/14732 [00:27<00:00, 898.62 examples/s]
Map: 100%  819/819 [00:01<00:00, 811.68 examples/s]
Map: 100%  818/818 [00:01<00:00, 564.12 examples/s]
Dataset({
    features: ['input_ids', 'attention_mask', 'labels'],
    num_rows: 14732
})
Dataset({
    features: ['input_ids', 'attention_mask', 'labels'],
    num_rows: 819
})
Dataset({
    features: ['input_ids', 'attention_mask', 'labels'],
    num_rows: 818
})
```

```
sample=tokenized_train[0]
print(sample['input_ids'])
print(sample['attention_mask'])
print(sample['labels'])
```

```
[0, 10127, 5219, 35, 38, 17241, 1437, 15269, 4, 1832, 47, 236, 103, 116, 50121, 50118, 39237, 35, 9136, 328, 50121, 50118, 10127, 5219,
[1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]
[0, 10127, 5219, 17241, 15269, 8, 40, 836, 6509, 103, 3859, 4, 2]
```

```
from transformers import pipeline

summarizer=pipeline('summarization', model=os.getenv('MODEL'))

news="summarize:Melbourne, Australia's cultural capital, pulsates with a vibrant energy. Grand Victorian architecture mingles with modern la

summarizer(news)
```

```
[{'summary_text': "Melbourne, Australia, is one of the world's most visited cities, according to Lonely Planet."}]
```

```
from transformers import BartForConditionalGeneration

model=BartForConditionalGeneration.from_pretrained(os.getenv('MODEL'))
print(model)
```

```
BartForConditionalGeneration(
  (model): BartModel(
    (shared): Embedding(50264, 1024, padding_idx=1)
    (encoder): BartEncoder(
      (embed_tokens): Embedding(50264, 1024, padding_idx=1)
      (embed_positions): BartLearnedPositionalEmbedding(1026, 1024)
      (layers): ModuleList(
        (0-11): 12 x BartEncoderLayer(
          (self_attn): BartSdpaAttention(
            (k_proj): Linear(in_features=1024, out_features=1024, bias=True)
            (v_proj): Linear(in_features=1024, out_features=1024, bias=True)
            (q_proj): Linear(in_features=1024, out_features=1024, bias=True)
            (out_proj): Linear(in_features=1024, out_features=1024, bias=True)
          )
          (self_attn_layer_norm): LayerNorm((1024,), eps=1e-05, elementwise_affine=True)
          (activation_fn): GELUActivation()
          (fc1): Linear(in_features=1024, out_features=4096, bias=True)
```

```
            (fc2): Linear(in_features=4096, out_features=1024, bias=True)
            (final_layer_norm): LayerNorm((1024,), eps=1e-05, elementwise_affine=True)
          )
        )
        (layernorm_embedding): LayerNorm((1024,), eps=1e-05, elementwise_affine=True)
      )
      (decoder): BartDecoder(
        (embed_tokens): Embedding(50264, 1024, padding_idx=1)
        (embed_positions): BartLearnedPositionalEmbedding(1026, 1024)
        (layers): ModuleList(
          (0-11): 12 x BartDecoderLayer(
            (self_attn): BartSdpaAttention(
              (k_proj): Linear(in_features=1024, out_features=1024, bias=True)
              (v_proj): Linear(in_features=1024, out_features=1024, bias=True)
              (q_proj): Linear(in_features=1024, out_features=1024, bias=True)
              (out_proj): Linear(in_features=1024, out_features=1024, bias=True)
            )
            (activation_fn): GELUActivation()
            (self_attn_layer_norm): LayerNorm((1024,), eps=1e-05, elementwise_affine=True)
            (encoder_attn): BartSdpaAttention(
              (k_proj): Linear(in_features=1024, out_features=1024, bias=True)
              (v_proj): Linear(in_features=1024, out_features=1024, bias=True)
              (q_proj): Linear(in_features=1024, out_features=1024, bias=True)
              (out_proj): Linear(in_features=1024, out_features=1024, bias=True)
            )
            (encoder_attn_layer_norm): LayerNorm((1024,), eps=1e-05, elementwise_affine=True)
            (fc1): Linear(in_features=1024, out_features=4096, bias=True)
            (fc2): Linear(in_features=4096, out_features=1024, bias=True)
            (final_layer_norm): LayerNorm((1024,), eps=1e-05, elementwise_affine=True)
          )
        )
        (layernorm_embedding): LayerNorm((1024,), eps=1e-05, elementwise_affine=True)
      )
    )
    (lm_head): Linear(in_features=1024, out_features=50264, bias=False)
  )
```

```
from transformers import DataCollatorForSeq2Seq

data_collator= DataCollatorForSeq2Seq(tokenizer=tokenizer, model=model)
print(data_collator)
```

```
  DataCollatorForSeq2Seq(tokenizer=BartTokenizer(name_or_path='facebook/bart-large-xsum', vocab_size=50265, model_max_length=1024, is_f
        0: AddedToken("<s>", rstrip=False, lstrip=False, single_word=False, normalized=True, special=True),
        1: AddedToken("<pad>", rstrip=False, lstrip=False, single_word=False, normalized=True, special=True),
        2: AddedToken("</s>", rstrip=False, lstrip=False, single_word=False, normalized=True, special=True),
        3: AddedToken("<unk>", rstrip=False, lstrip=False, single_word=False, normalized=True, special=True),
        50264: AddedToken("<mask>", rstrip=False, lstrip=True, single_word=False, normalized=True, special=True),
  }, model=BartForConditionalGeneration(
    (model): BartModel(
      (shared): Embedding(50264, 1024, padding_idx=1)
      (encoder): BartEncoder(
        (embed_tokens): Embedding(50264, 1024, padding_idx=1)
        (embed_positions): BartLearnedPositionalEmbedding(1026, 1024)
        (layers): ModuleList(
          (0-11): 12 x BartEncoderLayer(
            (self_attn): BartSdpaAttention(
              (k_proj): Linear(in_features=1024, out_features=1024, bias=True)
              (v_proj): Linear(in_features=1024, out_features=1024, bias=True)
              (q_proj): Linear(in_features=1024, out_features=1024, bias=True)
              (out_proj): Linear(in_features=1024, out_features=1024, bias=True)
            )
            (self_attn_layer_norm): LayerNorm((1024,), eps=1e-05, elementwise_affine=True)
            (activation_fn): GELUActivation()
            (fc1): Linear(in_features=1024, out_features=4096, bias=True)
            (fc2): Linear(in_features=4096, out_features=1024, bias=True)
            (final_layer_norm): LayerNorm((1024,), eps=1e-05, elementwise_affine=True)
          )
        )
        (layernorm_embedding): LayerNorm((1024,), eps=1e-05, elementwise_affine=True)
      )
      (decoder): BartDecoder(
        (embed_tokens): Embedding(50264, 1024, padding_idx=1)
        (embed_positions): BartLearnedPositionalEmbedding(1026, 1024)
        (layers): ModuleList(
          (0-11): 12 x BartDecoderLayer(
            (self_attn): BartSdpaAttention(
              (k_proj): Linear(in_features=1024, out_features=1024, bias=True)
              (v_proj): Linear(in_features=1024, out_features=1024, bias=True)
              (q_proj): Linear(in_features=1024, out_features=1024, bias=True)
              (out_proj): Linear(in_features=1024, out_features=1024, bias=True)
            )
            (activation_fn): GELUActivation()
```

```
            (self_attn_layer_norm): LayerNorm((1024,), eps=1e-05, elementwise_affine=True)
            (encoder_attn): BartSdpaAttention(
              (k_proj): Linear(in_features=1024, out_features=1024, bias=True)
              (v_proj): Linear(in_features=1024, out_features=1024, bias=True)
              (q_proj): Linear(in_features=1024, out_features=1024, bias=True)
              (out_proj): Linear(in_features=1024, out_features=1024, bias=True)
            )
            (encoder_attn_layer_norm): LayerNorm((1024,), eps=1e-05, elementwise_affine=True)
            (fc1): Linear(in_features=1024, out_features=4096, bias=True)
            (fc2): Linear(in_features=4096, out_features=1024, bias=True)
            (final_layer_norm): LayerNorm((1024,), eps=1e-05, elementwise_affine=True)
          )
        )
        (layernorm_embedding): LayerNorm((1024,), eps=1e-05, elementwise_affine=True)
      )
    )
```

```
from datasets import load_metric

metric=load_metric('rouge')
```

```
import nltk
import numpy as np

nltk.download('punkt')

def compute_metrics(eval_pred):
    predictions, labels=eval_pred

    decoded_preds=tokenizer.batch_decode(predictions, skip_special_tokens=True)

    labels=np.where(labels!=-100, labels, tokenizer.pad_token_id)
    decoded_labels=tokenizer.batch_decode(labels, skip_special_tokens=True)

    decoded_preds=['\n'.join(nltk.sent_tokenize(pred.strip())) for pred in decoded_preds]
    decoded_labels=['\n'.join(nltk.sent_tokenize(label.strip())) for label in decoded_labels]

    result=metric.compute(predictions=decoded_preds, references=decoded_labels, use_stemmer=True)
    result={key: value.mid.fmeasure*100 for key, value in result.items()}

    prediction_lens=[np.count_nonzero(pred!=tokenizer.pad_token_id) for pred in predictions]
    result['gen_len']=np.mean(prediction_lens)
    return {k: round(v,4) for k,v in result.items()}
```

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Package punkt is already up-to-date!
```

```
!pip install accelerate==0.21.0
```

```
Requirement already satisfied: accelerate==0.21.0 in /usr/local/lib/python3.10/dist-packages (0.21.0)
Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.10/dist-packages (from accelerate==0.21.0) (1.25.2)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist-packages (from accelerate==0.21.0) (24.0)
Requirement already satisfied: psutil in /usr/local/lib/python3.10/dist-packages (from accelerate==0.21.0) (5.9.5)
Requirement already satisfied: pyyaml in /usr/local/lib/python3.10/dist-packages (from accelerate==0.21.0) (6.0.1)
Requirement already satisfied: torch>=1.10.0 in /usr/local/lib/python3.10/dist-packages (from accelerate==0.21.0) (2.2.1+cu121)
Requirement already satisfied: filelock in /usr/local/lib/python3.10/dist-packages (from torch>=1.10.0->accelerate==0.21.0) (3.13.4)
Requirement already satisfied: typing-extensions>=4.8.0 in /usr/local/lib/python3.10/dist-packages (from torch>=1.10.0->accelerate==0.21
Requirement already satisfied: sympy in /usr/local/lib/python3.10/dist-packages (from torch>=1.10.0->accelerate==0.21.0) (1.12)
Requirement already satisfied: networkx in /usr/local/lib/python3.10/dist-packages (from torch>=1.10.0->accelerate==0.21.0) (3.3)
Requirement already satisfied: jinja2 in /usr/local/lib/python3.10/dist-packages (from torch>=1.10.0->accelerate==0.21.0) (3.1.3)
Requirement already satisfied: fsspec in /usr/local/lib/python3.10/dist-packages (from torch>=1.10.0->accelerate==0.21.0) (2023.6.0)
Requirement already satisfied: nvidia-cuda-nvrtc-cu12==12.1.105 in /usr/local/lib/python3.10/dist-packages (from torch>=1.10.0->accelera
Requirement already satisfied: nvidia-cuda-runtime-cu12==12.1.105 in /usr/local/lib/python3.10/dist-packages (from torch>=1.10.0->accele
Requirement already satisfied: nvidia-cuda-cupti-cu12==12.1.105 in /usr/local/lib/python3.10/dist-packages (from torch>=1.10.0->accelera
Requirement already satisfied: nvidia-cudnn-cu12==8.9.2.26 in /usr/local/lib/python3.10/dist-packages (from torch>=1.10.0->accelerate==0
Requirement already satisfied: nvidia-cublas-cu12==12.1.3.1 in /usr/local/lib/python3.10/dist-packages (from torch>=1.10.0->accelerate==
Requirement already satisfied: nvidia-cufft-cu12==11.0.2.54 in /usr/local/lib/python3.10/dist-packages (from torch>=1.10.0->accelerate==
Requirement already satisfied: nvidia-curand-cu12==10.3.2.106 in /usr/local/lib/python3.10/dist-packages (from torch>=1.10.0->accelerate
Requirement already satisfied: nvidia-cusolver-cu12==11.4.5.107 in /usr/local/lib/python3.10/dist-packages (from torch>=1.10.0->accelera
Requirement already satisfied: nvidia-cusparse-cu12==12.1.0.106 in /usr/local/lib/python3.10/dist-packages (from torch>=1.10.0->accelera
Requirement already satisfied: nvidia-nccl-cu12==2.19.3 in /usr/local/lib/python3.10/dist-packages (from torch>=1.10.0->accelerate==0.21
Requirement already satisfied: nvidia-nvtx-cu12==12.1.105 in /usr/local/lib/python3.10/dist-packages (from torch>=1.10.0->accelerate==0.
Requirement already satisfied: triton==2.2.0 in /usr/local/lib/python3.10/dist-packages (from torch>=1.10.0->accelerate==0.21.0) (2.2.0)
Requirement already satisfied: nvidia-nvjitlink-cu12 in /usr/local/lib/python3.10/dist-packages (from nvidia-cusolver-cu12==11.4.5.107->
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.10/dist-packages (from jinja2->torch>=1.10.0->accelerate==0.21.
Requirement already satisfied: mpmath>=0.19 in /usr/local/lib/python3.10/dist-packages (from sympy->torch>=1.10.0->accelerate==0.21.0) (
```

```
pip install wandb
```

```
Requirement already satisfied: wandb in /usr/local/lib/python3.10/dist-packages (0.16.6)
Requirement already satisfied: Click!=8.0.0,>=7.1 in /usr/local/lib/python3.10/dist-packages (from wandb) (8.1.7)
Requirement already satisfied: GitPython!=3.1.29,>=1.0.0 in /usr/local/lib/python3.10/dist-packages (from wandb) (3.1.43)
Requirement already satisfied: requests<3,>=2.0.0 in /usr/local/lib/python3.10/dist-packages (from wandb) (2.31.0)
Requirement already satisfied: psutil>=5.0.0 in /usr/local/lib/python3.10/dist-packages (from wandb) (5.9.5)
Requirement already satisfied: sentry-sdk>=1.0.0 in /usr/local/lib/python3.10/dist-packages (from wandb) (2.0.1)
Requirement already satisfied: docker-pycreds>=0.4.0 in /usr/local/lib/python3.10/dist-packages (from wandb) (0.4.0)
Requirement already satisfied: PyYAML in /usr/local/lib/python3.10/dist-packages (from wandb) (6.0.1)
Requirement already satisfied: setproctitle in /usr/local/lib/python3.10/dist-packages (from wandb) (1.3.3)
Requirement already satisfied: setuptools in /usr/local/lib/python3.10/dist-packages (from wandb) (67.7.2)
Requirement already satisfied: appdirs>=1.4.3 in /usr/local/lib/python3.10/dist-packages (from wandb) (1.4.4)
Requirement already satisfied: protobuf!=4.21.0,<5,>=3.19.0 in /usr/local/lib/python3.10/dist-packages (from wandb) (3.20.3)
Requirement already satisfied: six>=1.4.0 in /usr/local/lib/python3.10/dist-packages (from docker-pycreds>=0.4.0->wandb) (1.16.0)
Requirement already satisfied: gitdb<5,>=4.0.1 in /usr/local/lib/python3.10/dist-packages (from GitPython!=3.1.29,>=1.0.0->wandb) (4.0.1
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests<3,>=2.0.0->wandb) (3.3
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests<3,>=2.0.0->wandb) (3.7)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests<3,>=2.0.0->wandb) (2.0.7)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests<3,>=2.0.0->wandb) (2024.2.2)
Requirement already satisfied: smmap<6,>=3.0.1 in /usr/local/lib/python3.10/dist-packages (from gitdb<5,>=4.0.1->GitPython!=3.1.29,>=1.0
```

```
from transformers import Seq2SeqTrainingArguments, Seq2SeqTrainer

training_args=Seq2SeqTrainingArguments(
    output_dir=os.getenv('WANDB_NAME'),
    evaluation_strategy='epoch',
    save_strategy='epoch',
    load_best_model_at_end=True,
    metric_for_best_model='eval_loss',
    seed=42,
    learning_rate=2e-5,
    max_steps=100,
    per_device_train_batch_size=4,
    per_device_eval_batch_size=4,
    gradient_accumulation_steps=4,
    weight_decay=0.01,
    save_total_limit=2,
    num_train_epochs=10, # only for testing
    predict_with_generate=True,
    fp16=True,
    report_to='wandb',
    run_name=os.getenv('WANDB_NAME')
)

trainer=Seq2SeqTrainer(
    model=model,
    args=training_args,
    train_dataset=tokenized_train,
    eval_dataset=tokenized_test,
    tokenizer=tokenizer,
    data_collator=data_collator,
    compute_metrics=compute_metrics
)

trainer.train()
```

```
wandb: Currently logged in as: mrami010 (odu_ram). Use `wandb login --relogin` to force
Tracking run with wandb version 0.16.6
Run data is saved locally in /content/wandb/run-20240501_015042-mjhbvy1c
Syncing run ft-facebook-bart-large-xsum-on-samsum to Weights & Biases (docs)
View project at https://wandb.ai/odu_ram/Fine-tuning%20BART%20Series%20LLMs
View run at https://wandb.ai/odu_ram/Fine-tuning%20BART%20Series%20LLMs/runs/mjhbvy1c
```


[100/100 06:39, Epoch 0/1]

Epoch	Training Loss	Validation Loss	Rouge1	Rouge2	Rougel	Rougelsum	Gen Len
0	No log	1.551391	49.173800	23.682000	40.079300	44.838200	26.081800

```
Some non-default generation parameters are set in the model config. These should go into
Non-default generation parameters: {'max_length': 62, 'min_length': 11, 'early_stopping'
There were missing keys in the checkpoint model loaded: ['model.encoder.embed_tokens.wei
TrainOutput(global_step=100, training_loss=1.7061257934570313, metrics=
{'train_runtime': 407.2204, 'train_samples_per_second': 3.929,
```

```
validation=trainer.evaluate(eval_dataset=tokenized_val)
print(validation)
```

[205/205 04:41]

```
{'eval_loss': 1.5248808860778809, 'eval_rouge1': 50.3616, 'eval_rouge2': 25.1246, 'eval_
```

```
test=trainer.evaluate(eval_dataset=tokenized_test)
print(test)
```

[205/205 20:10]

```
{'eval_loss': 1.551391363143921, 'eval_rouge1': 49.1738, 'eval_rouge2': 23.682, 'eval_rc
```

```
from transformers import GenerationConfig
kwargs={
    'model_name': f'{os.getenv("WANDB_NAME")}',
    'finetuned_from': f'{os.getenv("MODEL")}',
    'tasks': 'summarization'
}

trainer.push_to_hub(**kwargs)
tokenizer.push_to_hub(os.getenv('WANDB_NAME'))

generation_config=GenerationConfig(
    max_length=62, min_length=11, early_stopping=True, num_beams=6, no_repeat_ngram_size=3, forced_eos_token_id=2
)

generation_config.save_pretrained('aisuko/'+os.getenv('WANDB_NAME'), push_to_hub=True)
```

```
Some non-default generation parameters are set in the model config. These should go into
Non-default generation parameters: {'max_length': 62, 'min_length': 11, 'early_stopping'
```

training_args.bin: 100% 4.92k/4.92k [00:00<00:00, 15.2kB/s]

Upload 2 LFS files: 100% 2/2 [00:44<00:00, 44.77s/it]

model.safetensors: 100% 1.63G/1.63G [00:44<00:00, 44.0MB/s]

README.md: 100% 1.71k/1.71k [00:00<00:00, 122kB/s]

```
print(val_ds[35]['dialogue'])
```

```
John: doing anything special?
Alex: watching 'Millionaires' on tvn
Sam: me too! He has a chance to win a million!
John: ok, fingers crossed then! :)
```

```
summarizer(val_ds[35]['dialogue'], max_length=48)[0]
```

```
{'summary_text': 'A few weeks ago, I was chatting to my friends John and Alex on Facebook.'}
```

```
!pip install openai
```

```
Collecting openai
  Downloading openai-1.23.6-py3-none-any.whl (311 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 311.6/311.6 kB 2.8 MB/s eta 0:00:00
Requirement already satisfied: anyio<5,>=3.5.0 in /usr/local/lib/python3.10/dist-packages (from openai) (3.7.1)
Requirement already satisfied: distro<2,>=1.7.0 in /usr/lib/python3/dist-packages (from openai) (1.7.0)
Collecting httpx<1,>=0.23.0 (from openai)
  Downloading httpx-0.27.0-py3-none-any.whl (75 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 75.6/75.6 kB 5.5 MB/s eta 0:00:00
Requirement already satisfied: pydantic<3,>=1.9.0 in /usr/local/lib/python3.10/dist-packages (from openai) (2.7.0)
Requirement already satisfied: sniffio in /usr/local/lib/python3.10/dist-packages (from openai) (1.3.1)
Requirement already satisfied: tqdm>4 in /usr/local/lib/python3.10/dist-packages (from openai) (4.66.2)
Requirement already satisfied: typing-extensions<5,>=4.7 in /usr/local/lib/python3.10/dist-packages (from openai) (4.11.0)
Requirement already satisfied: idna>=2.8 in /usr/local/lib/python3.10/dist-packages (from anyio<5,>=3.5.0->openai) (3.7)
Requirement already satisfied: exceptiongroup in /usr/local/lib/python3.10/dist-packages (from anyio<5,>=3.5.0->openai) (1.2.1)
Requirement already satisfied: certifi in /usr/local/lib/python3.10/dist-packages (from httpx<1,>=0.23.0->openai) (2024.2.2)
Collecting httpcore==1.* (from httpx<1,>=0.23.0->openai)
  Downloading httpcore-1.0.5-py3-none-any.whl (77 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 77.9/77.9 kB 4.1 MB/s eta 0:00:00
Collecting h11<0.15,>=0.13 (from httpcore==1.*->httpx<1,>=0.23.0->openai)
  Downloading h11-0.14.0-py3-none-any.whl (58 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 58.3/58.3 kB 5.9 MB/s eta 0:00:00
Requirement already satisfied: annotated-types>=0.4.0 in /usr/local/lib/python3.10/dist-packages (from pydantic<3,>=1.9.0->openai) (0.6.
Requirement already satisfied: pydantic-core==2.18.1 in /usr/local/lib/python3.10/dist-packages (from pydantic<3,>=1.9.0->openai) (2.18.
Installing collected packages: h11, httpcore, httpx, openai
Successfully installed h11-0.14.0 httpcore-1.0.5 httpx-0.27.0 openai-1.23.6
```

```
pip install datasets
```

```
Collecting datasets
  Downloading datasets-2.19.0-py3-none-any.whl (542 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 542.0/542.0 kB 5.3 MB/s eta 0:00:00
Requirement already satisfied: filelock in /usr/local/lib/python3.10/dist-packages (from datasets) (3.13.4)
Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.10/dist-packages (from datasets) (1.25.2)
Requirement already satisfied: pyarrow>=12.0.0 in /usr/local/lib/python3.10/dist-packages (from datasets) (14.0.2)
Requirement already satisfied: pyarrow-hotfix in /usr/local/lib/python3.10/dist-packages (from datasets) (0.6)
Collecting dill<0.3.9,>=0.3.0 (from datasets)
  Downloading dill-0.3.8-py3-none-any.whl (116 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 116.3/116.3 kB 6.8 MB/s eta 0:00:00
Requirement already satisfied: pandas in /usr/local/lib/python3.10/dist-packages (from datasets) (2.0.3)
Requirement already satisfied: requests>=2.19.0 in /usr/local/lib/python3.10/dist-packages (from datasets) (2.31.0)
Requirement already satisfied: tqdm>=4.62.1 in /usr/local/lib/python3.10/dist-packages (from datasets) (4.66.2)
Collecting xxhash (from datasets)
  Downloading xxhash-3.4.1-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (194 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 194.1/194.1 kB 7.9 MB/s eta 0:00:00
Collecting multiprocess (from datasets)
  Downloading multiprocess-0.70.16-py310-none-any.whl (134 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 134.8/134.8 kB 7.9 MB/s eta 0:00:00
Requirement already satisfied: fsspec[http]<=2024.3.1,>=2023.1.0 in /usr/local/lib/python3.10/dist-packages (from datasets) (2023.6.0)
Requirement already satisfied: aiohttp in /usr/local/lib/python3.10/dist-packages (from datasets) (3.9.5)
Collecting huggingface-hub>=0.21.2 (from datasets)
  Downloading huggingface_hub-0.22.2-py3-none-any.whl (388 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 388.9/388.9 kB 10.4 MB/s eta 0:00:00
Requirement already satisfied: packaging in /usr/local/lib/python3.10/dist-packages (from datasets) (24.0)
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.10/dist-packages (from datasets) (6.0.1)
Requirement already satisfied: aiosignal>=1.1.2 in /usr/local/lib/python3.10/dist-packages (from aiohttp->datasets) (1.3.1)
Requirement already satisfied: attrs>=17.3.0 in /usr/local/lib/python3.10/dist-packages (from aiohttp->datasets) (23.2.0)
Requirement already satisfied: frozenlist>=1.1.1 in /usr/local/lib/python3.10/dist-packages (from aiohttp->datasets) (1.4.1)
Requirement already satisfied: multidict<7.0,>=4.5 in /usr/local/lib/python3.10/dist-packages (from aiohttp->datasets) (6.0.5)
Requirement already satisfied: yarl<2.0,>=1.0 in /usr/local/lib/python3.10/dist-packages (from aiohttp->datasets) (1.9.4)
Requirement already satisfied: async-timeout<5.0,>=4.0 in /usr/local/lib/python3.10/dist-packages (from aiohttp->datasets) (4.0.3)
Requirement already satisfied: typing-extensions>=3.7.4.3 in /usr/local/lib/python3.10/dist-packages (from huggingface-hub>=0.21.2->data
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests>=2.19.0->datasets) (3.
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests>=2.19.0->datasets) (3.7)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests>=2.19.0->datasets) (2.0.7)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests>=2.19.0->datasets) (2024.2.2
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.10/dist-packages (from pandas->datasets) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-packages (from pandas->datasets) (2023.4)
Requirement already satisfied: tzdata>=2022.1 in /usr/local/lib/python3.10/dist-packages (from pandas->datasets) (2024.1)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.8.2->pandas->datasets) (1.16
Installing collected packages: xxhash, dill, multiprocess, huggingface-hub, datasets
  Attempting uninstall: huggingface-hub
    Found existing installation: huggingface-hub 0.20.3
    Uninstalling huggingface-hub-0.20.3:
      Successfully uninstalled huggingface-hub-0.20.3
Successfully installed datasets-2.19.0 dill-0.3.8 huggingface-hub-0.22.2 multiprocess-0.70.16 xxhash-3.4.1
```

```
from transformers import pipeline, set_seed

import matplotlib.pyplot as plt

from datasets import load_dataset, load_metric

import pandas as pd

from transformers import AutoModelForSeq2SeqLM, AutoTokenizer

import nltk

from nltk.tokenize import sent_tokenize

from tqdm import tqdm

import torch

nltk.download("punkt")
```

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Unzipping tokenizers/punkt.zip.
True
```

```
from  openai import OpenAI
```

```
dataset_samsum = load_dataset("samsum")
```

```
/usr/local/lib/python3.10/dist-packages/huggingface_hub/utils/_token.py:89: UserWarning:
The secret `HF_TOKEN` does not exist in your Colab secrets.
To authenticate with the Hugging Face Hub, create a token in your settings tab (https://
You will be able to reuse this secret in all of your notebooks.
Please note that authentication is recommended but still optional to access public model
  warnings.warn(
Downloading data: 100% 6.06M/6.06M [00:00<00:00, 17.7MB/s]
Downloading data: 100% 347k/347k [00:00<00:00, 1.37MB/s]
Downloading data: 100% 335k/335k [00:00<00:00, 1.06MB/s]
Generating train split: 100% 14732/14732 [00:00<00:00, 62197.15 examples/s]
Generating test split: 100% 819/819 [00:00<00:00, 22290.15 examples/s]
Generating validation split: 100% 818/818 [00:00<00:00, 23769.03 examples/s]
DatasetDict({
    train: Dataset({
        features: ['id', 'dialogue', 'summary'],
        num_rows: 14732
    })
    test: Dataset({
        features: ['id', 'dialogue', 'summary'],
        num_rows: 819
    })
    validation: Dataset({
        features: ['id', 'dialogue', 'summary'],
        num_rows: 818
    })
})
```

```
few_shotexample = []
for i in range(50):
  di=dataset_samsum['train'][i]
  di.pop('id')
  value=di.pop('dialogue')
  di['sentence']=value
  value=di.pop('summary')
  di['summary']=value
  few_shotexample.append(di)
print(few_shotexample)
```

```
[{'sentence': "Amanda: I baked  cookies. Do you want some?\r\nJerry: Sure!\r\nAmanda: I'll bring you tomorrow :-)", 'summary': 'Amanda b
```

```
"""### workon o dartaset it has to be in below format

few_shotexample = [{

    "sentence":"Amanda: I baked cookies. Do you want some? Jerry: Sure! Amanda: I'll bring you tomorrow :-)",
    "summary":"Amanda baked cookies and will bring Jerry some tomorrow."

},
{

   "sentence":"Olivia: Who are you voting for in this election? Oliver: Liberals as always. Olivia: Me too!! Oliver: Great",
    "summary":"Olivia and Olivier are voting for liberals in this election."

},
#{

    #"sentence":"Tim: Hi, what's up? Kim: Bad mood tbh, I was going to do lots of stuff but ended up procrastinating Tim: What did you plan
   # "summary":"Kim may try the pomodoro technique recommended by Tim to get more stuff done"

#}

]"""
```

```
client = OpenAI(api_key="sk-proj-fje2CNVx5hBAegDzAE9KT3BlbkFJyhwZRkWL1oktliWnCUjN")

def get_completion(prompt, model="gpt-3.5-turbo"):
    messages = [{"role":'system', 'content': "Consider You are a expert  model who can summarize  dailagues from the give sentences.",
                 "role":"system", "content":f""" You are given some examples which is in json format, where first key is sentence and other
        ```{few_shotexample}```""",
        "role": "user", "content": prompt}]
    response = client.chat.completions.create(
        model=model,
        messages=messages,
        temperature=0,
    )
    return response
```

```
summary = []
for i in range(10, 70):
    di = dataset_samsum['train'][i]
    summary = di.pop('summary')
    sentence.append([summary])

print(sentence)
```

```
[['Demi got promoted. She will celebrate that with Lucas at Death & Co at 10 pm.'], ['Mark just shipped the goods and he will send Georg
```

```
prediction_result=[]
for i in range(10,70):
  di=dataset_samsum['train'][i]
  di.pop('id')
  value=di.pop('dialogue')
  di['sentence']=value
  value=di.pop('summary')
  di['summary']=value
  message = f"""Based on the given sentence  I want you to act as an expert who can summarize the whole dailogues with shorter length than se
  The sentences is ;;;{di["sentence"]};;;"""
  completion = get_completion(message, model="gpt-3.5-turbo-1106")
  prediction_result.append(completion.choices[0].message.content)
  print(completion.choices[0].message.content, "\n")
```

```
Lucas and Demi have a brief conversation where Demi shares that she had a good day and got promoted. Lucas congratulates her and they

Mark informed George that he has shipped the goods and will send the tracking number tomorrow. George expressed his gratitude.

Anita is at the Bologna station and everything is going smoothly.

Leon asked Arthur if he found a job yet, and Arthur said he was still unemployed. Leon then mentioned a job opportunity as a junior p

Macca was excited to try ice climbing for the first time near Reykjavik. Despite some initial fear, he enjoyed the experience and fou

Isabella is regretting her actions at the Christmas party and is considering calling in sick to work. Oscar is teasing her about the

Tina reminded Lucy about owing her 50 bucks. Lucy confirmed and said she already transferred the money, but it will be in Tina's acco
```
```

```
Betty regrets drinking too much wine and is embarrassed about her behavior while drunk. Amber teases her about being so drunk that sh
Mary and Mike greet each other. Mike is visiting his grandma and invites Mary to come along because his grandma likes her. Mary agree
Laura finished her work for the day and agreed to wait for Kim at work so they could go home together. Kim suggested meeting at 7 and
Ashley recommends a life-changing book to her friends, but they have different opinions on it. Marcus and Seamus prefer books that ar
Aria met Charlie Evans, who is doing great, married, runs a family business, and has two daughters. Maverick reminisces about old tim
Anna asks Omenah where she is, and Omenah responds that she is at home. Anna then says she will be there in a minute.
Renee and Rachel catch up, with Renee sharing that Layla is having knee surgery due to arthritis. Rachel is settling in and looking f
Jonas is running late and asks his colleagues to let Mary know he will present today. Natalie agrees and Olivia saves a seat for him.
Julius and Lawrence are both disappointed with Manchester United's performance this season. They feel that the team has quality but n
Jade and Wayne discuss a postponed trip, with Jade expressing surprise and disappointment, and Wayne reassuring her that she can stil
Natalie asked Jason if he was still going to Thailand and requested him to buy her some spices. Jason agreed and asked for the names
Elisa is inviting everyone for drinks at Mombasa tonight. Alice, Sadie, Carol, Arthur, Liam, Kai, Tom, and John are all in. Elisa is
Hal asks Amy about her homework, Amy confirms she has none. Hal reminds her that her mom is not home and Amy says she can use the mic
Ray's bike was stolen and he's asking for help to share it on Facebook. Others share their own experiences with stolen bikes and offe
Eric and Curtis discuss the upcoming Champions League. Eric mentions the date and Curtis talks about predictions. Eric admits to losi
Gunther asked if Chandler paid for coffee, and Chandler admitted he didn't but said he would pay tomorrow. Gunther looked unimpressed
Karen's friends complimented her on a photo, and they all wanted to know where she got her dress. Karen thanked them and told them it
The group plans a weekend reunion, with everyone agreeing to meet at the same place as last time on Friday night. Ted will book the r
Bradley and Julianna discuss safety in Europe and the strict laws regarding touching children. They also mention the presence of pedo
Lucia needs a haircut because she is changing jobs and her hair must be shorter. She schedules an appointment with Eric at his beauty
Gabriella asked Jasmine to review her CV in English for a job application, and Jasmine fixed a few grammar and spelling mistakes. Gab
```

```
from transformers import pipeline
from datasets import load_dataset
from rouge import Rouge
```

```
pip install rouge_score
```

```
Requirement already satisfied: rouge_score in /usr/local/lib/python3.10/dist-packages (0.1.2)
Requirement already satisfied: absl-py in /usr/local/lib/python3.10/dist-packages (from rouge_score) (1.4.0)
Requirement already satisfied: nltk in /usr/local/lib/python3.10/dist-packages (from rouge_score) (3.8.1)
Requirement already satisfied: numpy in /usr/local/lib/python3.10/dist-packages (from rouge_score) (1.25.2)
Requirement already satisfied: six>=1.14.0 in /usr/local/lib/python3.10/dist-packages (from rouge_score) (1.16.0)
Requirement already satisfied: click in /usr/local/lib/python3.10/dist-packages (from nltk->rouge_score) (8.1.7)
Requirement already satisfied: joblib in /usr/local/lib/python3.10/dist-packages (from nltk->rouge_score) (1.4.0)
Requirement already satisfied: regex>=2021.8.3 in /usr/local/lib/python3.10/dist-packages (from nltk->rouge_score) (2023.12.25)
Requirement already satisfied: tqdm in /usr/local/lib/python3.10/dist-packages (from nltk->rouge_score) (4.66.2)
```

```
summary1 = []
for i in range(10, 70):
    di = dataset_samsum['train'][i]
    summary = di.pop('summary')
    summary1.append([summary])
```

```
from datasets import load_metric

bleu_metric = load_metric('bleu')
rouge_metric = load_metric('rouge')

predictions= prediction_result
references = summary1

rouge_score = rouge_metric.compute(predictions=predictions, references=references)
print(f"ROUGE: {rouge_score}")
```

```
/usr/local/lib/python3.10/dist-packages/datasets/load.py:759: FutureWarning: The repository for bleu contains custom code which must be
You can avoid this message in future by passing the argument `trust_remote_code=True`.
Passing `trust_remote_code=True` will be mandatory to load this metric from the next major release of `datasets`.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/datasets/load.py:759: FutureWarning: The repository for rouge contains custom code which must be
You can avoid this message in future by passing the argument `trust_remote_code=True`.
Passing `trust_remote_code=True` will be mandatory to load this metric from the next major release of `datasets`.
  warnings.warn(
ROUGE: {'rouge1': AggregateScore(low=Score(precision=0.29312231319230153, recall=0.5283583142002882, fmeasure=0.3628884970678932), mid=S
```

```
predictions
```

```
['Lucas and Demi have a brief conversation where Demi shares that she had a good day and got promoted. Lucas congratulates her and they
make plans to celebrate at Death & Co. at 10pm.',
 'Mark informed George that he has shipped the goods and will send the tracking number tomorrow. George expressed his gratitude.',
 'Anita is at the Bologna station and everything is going smoothly.',
 "Leon asked Arthur if he found a job yet, and Arthur said he was still unemployed. Leon then mentioned a job opportunity as a junior
project manager at his friend's company and offered to help Arthur apply for it. Arthur was interested and asked for details, and Leon
provided his friend's email for Arthur to send his resume.",
 'Macca was excited to try ice climbing for the first time near Reykjavik in Iceland. Despite some initial fear, he enjoyed the
experience and found the landscapes magnificent.',
 "Isabella is regretting her actions at the Christmas party and doesn't want to go to work. Oscar is teasing her about it and they are
both joking about the events of the party. Isabella is upset with Oscar for not stopping her from drinking too much. They both think
they will be in trouble at work.",
 "Tina reminded Lucy about owing her 50 bucks. Lucy confirmed and said she already transferred the money, but it will be in Tina's
account tomorrow. Tina was relieved because she has been having a lot of expenses lately.",
 'Betty regrets drinking too much wine and is embarrassed about her behavior while drunk. Amber teases her about being drunk and doing
silly things.',
 "Mary and Mike greet each other. Mike is going to visit his grandma and invites Mary to come along. Mary agrees and plans to buy
chocolate for Mike's grandma.",
 'Laura finished her work for the day and agreed to wait for Kim at work so they could go home together. Kim suggested meeting at 7 and
Laura agreed.']
```

```
pip install sacrebleu
```

```
Collecting sacrebleu
  Downloading sacrebleu-2.4.2-py3-none-any.whl (106 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 106.7/106.7 kB 3.4 MB/s eta 0:00:00
Collecting portalocker (from sacrebleu)
  Downloading portalocker-2.8.2-py3-none-any.whl (17 kB)
Requirement already satisfied: regex in /usr/local/lib/python3.10/dist-packages (from sacrebleu) (2023.12.25)
Requirement already satisfied: tabulate>=0.8.9 in /usr/local/lib/python3.10/dist-packages (from sacrebleu) (0.9.0)
Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.10/dist-packages (from sacrebleu) (1.25.2)
Collecting colorama (from sacrebleu)
  Downloading colorama-0.4.6-py2.py3-none-any.whl (25 kB)
Requirement already satisfied: lxml in /usr/local/lib/python3.10/dist-packages (from sacrebleu) (4.9.4)
Installing collected packages: portalocker, colorama, sacrebleu
Successfully installed colorama-0.4.6 portalocker-2.8.2 sacrebleu-2.4.2
```

```
from datasets import load_metric
pedictions=prediction_result
references=summary1

# Load the SACREBLEU metric
sacrebleu_metric = load_metric("sacrebleu")
results = sacrebleu_metric.compute(predictions=predictions, references=references)

# Print the keys and the SACREBLEU score
print(list(results.keys()))
print(results["score"])
```

```
/usr/local/lib/python3.10/dist-packages/datasets/load.py:759: FutureWarning: The repository for sacrebleu contains custom code which mus
You can avoid this message in future by passing the argument `trust_remote_code=True`.
Passing `trust_remote_code=True` will be mandatory to load this metric from the next major release of `datasets`.
  warnings.warn(
['score', 'counts', 'totals', 'precisions', 'bp', 'sys_len', 'ref_len']
9.404998926301946
```