ShritejShrikant_Chavan_HW6d

November 4, 2023

[2]:

```
!nvidia-smi
```

```
Sat Nov  4 02:17:27 2023
+-----------------------------------------------------------------------------+
| NVIDIA-SMI 525.105.17   Driver Version: 525.105.17   CUDA Version: 12.0     |
|-------------------------------+----------------------+----------------------+
| 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 V100-SXM2…  Off  | 00000000:00:04.0 Off |                    0 |
| N/A   38C    P0    25W / 300W |      0MiB / 16384MiB |      0%      Default |
|                               |                      |                  N/A |
+-------------------------------+----------------------+----------------------+

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

0.1 Setup Environment and Install Libraries

[3]:

```
# CHANGE FOLDERS AS PER YOUR SETUP
from pathlib import Path
if 'google.colab' in str(get_ipython()):
    from google.colab import drive
    drive.mount("/content/drive")
    !pip install datasets transformers evaluate wandb accelerate seqeval -U -qq
    !pip install datasets transformers[sentencepiece] evaluate wandb accelerate␣
 ↪-U -qq
    !pip install sacrebleu bert_score -U -qq
    base_folder = Path("/content/drive/MyDrive/NLP")
else:
    base_folder = Path("/home/harpreet/Insync/google_drive_shaannoor/data")
```

```
from transformers import AutoConfig, AutoModelForSeq2SeqLM, AutoTokenizer,␣
 ↪Seq2SeqTrainer
from transformers import MarianMTModel, MarianTokenizer
from transformers import AutoTokenizer, DataCollatorForSeq2Seq, pipeline
from datasets import load_dataset, DatasetDict
import evaluate
from evaluate import evaluator

from transformers import GenerationConfig


from datasets import load_dataset, DatasetDict, Dataset, ClassLabel, Sequence
import evaluate


import wandb

import numpy as np
from sklearn.metrics import ConfusionMatrixDisplay
from sklearn.metrics import confusion_matrix
import seaborn as sns
import matplotlib.pyplot as plt
import pandas as pd
import torch
import gc

import textwrap
```

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call
drive.mount("/content/drive", force_remount=True).
```

0.2 Specify Model & Data Path

[4]:
```
# CHANGE FOLDERS TO WHERE YOU WANT TO SAVE DATA AND MODELS
data_folder = base_folder/'datasets/brown_corpus'
model_folder = base_folder/'models/nlp_spring_2023/kde4'
model_folder.mkdir(exist_ok=True)
data_folder.mkdir(exist_ok=True)
```

0.3 Create Functions

[5]:
```
def print_wrap(text, d):
    # If the text is a list, convert it to a string
    if isinstance(text, list):
        # Convert None values to a default string (e.g., "None" or an empty␣
↪string)
        text = ' '.join(str(item) if item is not None else "None" for item in␣
↪text)

    # Wrap the text to limit the width to 'd'
    wrapped_text = textwrap.fill(text, width=d)

    # Print the wrapped text
    print(wrapped_text)
```

0.3.1 Load and Split Dataset

```
[6]: def split_dataset(data = 'kde4', train = 1000, val_test = 500):
         kde_dataset = load_dataset('kde4', lang1='en', lang2='fr')

         test_val_splits = kde_dataset['train'].train_test_split(test_size=0.4,␣
     ↪seed=42)
         train_split= test_val_splits['train']
         test_val_splits = test_val_splits['test'].train_test_split(test_size=0.5,␣
     ↪seed=42,)
         val_split = test_val_splits['train']
         test_split = test_val_splits['test']



         train_split_small = train_split.shuffle(seed=42).select(range(train))
         val_split_small = val_split.shuffle(seed=42).select(range(val_test))
         test_split_small = test_split.shuffle(seed=42).select(range(val_test))

         # combine train, val splits into one dataset
         train_val_subset = DatasetDict({'train': train_split_small, 'val':␣
     ↪val_split_small})

         # create test dataset from test split
         test_subset = DatasetDict({'test': test_split_small})


         return train_val_subset, test_subset
```

0.3.2 Create function for Tokenizer

0.3.3 Function for Tokenization

[7]:
```python
def get_tokenized_dataset(checkpoint, dataset):
    tokenizer = AutoTokenizer.from_pretrained(checkpoint)


    max_length = 128
    def tokenize_fn(batch):
        inputs = [example['en'] for example in batch['translation']]
        targets = [example['fr'] for example in batch['translation']]
        model_inputs = tokenizer(text = inputs, text_target=targets, truncation = True, max_length=max_length)
        return model_inputs
        # CODE HERE

    tokenized_dataset = dataset.map(tokenize_fn,
                                    batched = True,
                                    remove_columns=dataset['train'].column_names)


    tokenized_dataset.set_format(type='torch')

    return tokenized_dataset
```

0.3.4 Function to initialize model

[8]:
```python
def initialize_model(checkpoint):

    config = AutoConfig.from_pretrained(checkpoint)

    generation_config = GenerationConfig.from_pretrained(checkpoint)

    model = AutoModelForSeq2SeqLM.from_pretrained(
    checkpoint,
    config=config,
    )

    return model, config
```

0.3.5 Function to compute metrics

[9]:
```python
bleu_metric = evaluate.load("sacrebleu")
bert_metric = evaluate.load('bertscore')

def compute_metrics(preds_and_labels):
    # preds are not logits but token ids
    # api is inconsistent here
    # we are not simply using argmax bu use 'beam search'
    preds, labels = preds_and_labels

    # convert predictions into words
    decoded_preds = tokenizer.batch_decode(preds, skip_special_tokens=True)

    # for any -100 label, replace with pad token id
    labels = np.where( labels != -100, labels, tokenizer.pad_token_id )

    # convert labels into words
    decoded_labels = tokenizer.batch_decode(labels, skip_special_tokens= True)

    # get rid of extra whitespace
    # and also, put targets into lists

    decoded_preds_cleaned = [pred.strip() for pred in decoded_preds]
    decoded_labels_cleaned = [label.strip() for label in decoded_labels]

    bleu_score = bleu_metric.compute(predictions=decoded_preds_cleaned,␣
↪references=decoded_labels_cleaned)
    bert_score = bert_metric.compute(predictions=decoded_preds_cleaned,␣
↪references=decoded_labels_cleaned, lang='fr')

    return{'bleu_score:': bleu_score['score'], 'bert_score': np.
↪mean(bert_score['f1'])}
    # CODE HERE
```

0.3.6 Function to set Trainer

[10]:
```
def get_trainer(model, training_args, tokenized_dataset, compute_metrics,␣
↪tokenizer, data_collator):
    trainer = Seq2SeqTrainer(
    model=model,
    args=training_args,
    train_dataset=tokenized_dataset["train"],
    eval_dataset=tokenized_dataset["val"],
    compute_metrics=compute_metrics,
    tokenizer=tokenizer,
    data_collator=data_collator,
    )
    return trainer
```

0.3.7 Plot Confusion Matrix

[11]:

```
def free_memory():
    """
    Attempts to free up memory by deleting variables and running Python's␣
 ↪garbage collector.
    """
    gc.collect()
    for device_id in range(torch.cuda.device_count()):
        torch.cuda.set_device(device_id)
        torch.cuda.empty_cache()
    gc.collect()
```

0.3.8 Function to tokenize dataset and, train and eval models

[26]:

```
def tokenize_train_evaluate_log(training_args, checkpoint, base_folder,␣
 ↪train_val_subset, compute_metrics):
    # 1. Free memory
    free_memory()

    # 2. Setup wandb
    wandb.login()
    %env WANDB_PROJECT = nlp_course_fall_2023-HW6-PartD

    ######################## ALLOWED TO CHANGE THIS BLOCK␣
 ↪################################################

    # MAKE SURE THE BASE FOLDER IS SETUP CORRECTLY
    #  YOU CAN CHANGE THIS LINE IF YOU WANT TO SAVE IN A DIFFERENT FOLDER

    model_folder = base_folder / "models" / "nlp_spring_2023/ner"/checkpoint
    model_folder.mkdir(exist_ok=True, parents=True)

    ######################## ALLOWED TO CHANGE THIS BLOCK␣
 ↪################################################



    # 3. Get Tokenized Dataset and Data Collator
    train_val_tokenized_dataset = get_tokenized_dataset(checkpoint,␣
 ↪train_val_subset)

    # 4. Initialize Model and Tokenizer
    model, config = initialize_model(checkpoint)
    tokenizer = AutoTokenizer.from_pretrained(checkpoint)

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

    trainer = get_trainer(model, training_args, train_val_tokenized_dataset,␣
↪compute_metrics, tokenizer, data_collator)

    # 6. Train and Evaluate
    trainer.train()
    trainer.evaluate(train_val_tokenized_dataset['val'])



    best_model_checkpoint_step = trainer.state.best_model_checkpoint.
↪split('-')[-1]
    wandb.log({"best_model_checkpoint_step": best_model_checkpoint_step})
    print(f"The best model was saved at step {best_model_checkpoint_step}.")

    wandb.finish()

    return best_model_checkpoint_step
```

0.4 Experiments

0.4.1 Experiment 1 - T5 Base model with 4 epochs

[27]:

```
from transformers import Seq2SeqTrainingArguments

# Define the directory where model checkpoints will be saved
model_folder = base_folder / "models" / "nlp_spring_2023/kde4/opus-mt-en-fr"

# Create the directory if it doesn't exist
model_folder.mkdir(exist_ok=True, parents=True)


checkpoint = 't5-base'
generation_config = GenerationConfig.from_pretrained(checkpoint)
tokenizer = AutoTokenizer.from_pretrained(checkpoint)




# Configure training parameters
training_args = Seq2SeqTrainingArguments(
    # Training-specific configurations
    num_train_epochs=4,  # Total number of training epochs
    weight_decay=0.01,  # Apply L2 regularization to prevent overfitting
    learning_rate=5e-5,  # Step size for the optimizer during training
    optim="adamw_torch",  # Optimizer,
    warmup_steps=10,
    predict_with_generate=True,
    #generation_config=generation_config,
    # memory and speed related arguments
    # Number of samples per training batch for each device
    per_device_train_batch_size=16,
    per_device_eval_batch_size=16,  # Number of samples per eval batch for each device

    gradient_checkpointing=True,  # memory
    # fp16 = True, # Speed
    # bf16=True,
    # tf32=True, # speed
    # evaluation settings
    output_dir=str(model_folder),  # Directory to save model checkpoints
    evaluation_strategy="steps",  # Evaluate model at specified step intervals
    eval_steps=10,  # Perform evaluation every 10 training steps
    # Checkpoint settings
    save_strategy="steps",  # Save model checkpoint at specified step intervals
    save_steps=10,  # Save a model checkpoint every 10 training steps
    load_best_model_at_end=True,  # Reload the best model at the end of training
    save_total_limit=2,  # Retain only the best and the most recent model checkpoints
    # metric_for_best_model=,
    # greater_is_better=,
    # Experiment logging configurations (commented out in this example)
    logging_strategy="steps",
    logging_steps=10,
    report_to="wandb",  # Log metrics and results to Weights & Biases platform
    # Experiment name for Weights & Biases
    run_name="translation-exp1",
)
```

```
/usr/local/lib/python3.10/dist-
packages/transformers/models/t5/tokenization_t5_fast.py:160: FutureWarning: This
tokenizer was incorrectly instantiated with a model max length of 512 which will
be corrected in Transformers v5.
For now, this behavior is kept to avoid breaking backwards compatibility when
padding/encoding with `truncation is True`.
- Be aware that you SHOULD NOT rely on t5-base automatically truncating your
input to 512 when padding/encoding.
- If you want to encode/pad to sequences longer than 512 you can either
instantiate this tokenizer with `model_max_length` or pass `max_length` when
encoding/padding.
- To avoid this warning, please instantiate this tokenizer with
`model_max_length` set to your preferred value.
  warnings.warn(
```

[28]:

```
train_val_subset, test_subset = split_dataset()
```

[29]:

```
training_args_dict = training_args.to_dict() # Convert TrainingArguments to␣
 ↪dictionary
#training_args_dict['run_name'] = f'{checkpoint}' # Update the run_name
new_training_args = Seq2SeqTrainingArguments(**training_args_dict)
```

```
/usr/local/lib/python3.10/dist-packages/transformers/training_args.py:1697:
FutureWarning: `--push_to_hub_token` is deprecated and will be removed in
version 5 of  Transformers. Use `--hub_token` instead.
  warnings.warn(
```

[30]:

```
best_model =tokenize_train_evaluate_log(training_args= new_training_args,␣
 ↪checkpoint=checkpoint, base_folder=base_folder,␣
 ↪train_val_subset=train_val_subset,
                             compute_metrics=compute_metrics)
```

```
env: WANDB_PROJECT=nlp_course_fall_2023-HW6-PartD
```

```
Map:   0%|          | 0/500 [00:00<?, ? examples/s]
```

```
<IPython.core.display.HTML object>
```

```
<IPython.core.display.HTML object>
```

```
<IPython.core.display.HTML object>
```

```
<IPython.core.display.HTML object>
```

```
<IPython.core.display.HTML object>
```

```
<IPython.core.display.HTML object>
```

```
You're using a T5TokenizerFast tokenizer. Please note that with a fast
tokenizer, using the `__call__` method is faster than using a method to encode
the text followed by a call to the `pad` method to get a padded encoding.
/usr/local/lib/python3.10/dist-packages/torch/utils/checkpoint.py:429:
UserWarning: torch.utils.checkpoint: please pass in use_reentrant=True or
use_reentrant=False explicitly. The default value of use_reentrant will be
updated to be False in the future. To maintain current behavior, pass
use_reentrant=True. It is recommended that you use use_reentrant=False. Refer to
docs for more details on the differences between the two variants.
  warnings.warn(
```

```
<IPython.core.display.HTML object>
```

```
/usr/local/lib/python3.10/dist-packages/transformers/generation/utils.py:1273:
UserWarning: Using the model-agnostic default `max_length` (=20) to control the
generation length. We recommend setting `max_new_tokens` to control the maximum
length of the generation.
  warnings.warn(
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
```

```
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
/usr/local/lib/python3.10/dist-packages/torch/utils/checkpoint.py:429:
UserWarning: torch.utils.checkpoint: please pass in use_reentrant=True or
use_reentrant=False explicitly. The default value of use_reentrant will be
updated to be False in the future. To maintain current behavior, pass
use_reentrant=True. It is recommended that you use use_reentrant=False. Refer to
docs for more details on the differences between the two variants.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/transformers/generation/utils.py:1273:
UserWarning: Using the model-agnostic default `max_length` (=20) to control the
generation length. We recommend setting `max_new_tokens` to control the maximum
length of the generation.
  warnings.warn(
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
/usr/local/lib/python3.10/dist-packages/torch/utils/checkpoint.py:429:
UserWarning: torch.utils.checkpoint: please pass in use_reentrant=True or
use_reentrant=False explicitly. The default value of use_reentrant will be
updated to be False in the future. To maintain current behavior, pass
use_reentrant=True. It is recommended that you use use_reentrant=False. Refer to
docs for more details on the differences between the two variants.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/transformers/generation/utils.py:1273:
UserWarning: Using the model-agnostic default `max_length` (=20) to control the
generation length. We recommend setting `max_new_tokens` to control the maximum
length of the generation.
  warnings.warn(
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
/usr/local/lib/python3.10/dist-packages/torch/utils/checkpoint.py:429:
UserWarning: torch.utils.checkpoint: please pass in use_reentrant=True or
use_reentrant=False explicitly. The default value of use_reentrant will be
updated to be False in the future. To maintain current behavior, pass
```

```
use_reentrant=True. It is recommended that you use use_reentrant=False. Refer to
docs for more details on the differences between the two variants.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/transformers/generation/utils.py:1273:
UserWarning: Using the model-agnostic default `max_length` (=20) to control the
generation length. We recommend setting `max_new_tokens` to control the maximum
length of the generation.
  warnings.warn(
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
/usr/local/lib/python3.10/dist-packages/torch/utils/checkpoint.py:429:
UserWarning: torch.utils.checkpoint: please pass in use_reentrant=True or
use_reentrant=False explicitly. The default value of use_reentrant will be
updated to be False in the future. To maintain current behavior, pass
use_reentrant=True. It is recommended that you use use_reentrant=False. Refer to
docs for more details on the differences between the two variants.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/transformers/generation/utils.py:1273:
UserWarning: Using the model-agnostic default `max_length` (=20) to control the
generation length. We recommend setting `max_new_tokens` to control the maximum
length of the generation.
  warnings.warn(
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
/usr/local/lib/python3.10/dist-packages/torch/utils/checkpoint.py:429:
UserWarning: torch.utils.checkpoint: please pass in use_reentrant=True or
use_reentrant=False explicitly. The default value of use_reentrant will be
updated to be False in the future. To maintain current behavior, pass
use_reentrant=True. It is recommended that you use use_reentrant=False. Refer to
docs for more details on the differences between the two variants.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/transformers/generation/utils.py:1273:
UserWarning: Using the model-agnostic default `max_length` (=20) to control the
generation length. We recommend setting `max_new_tokens` to control the maximum
length of the generation.
  warnings.warn(
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
```

```
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
/usr/local/lib/python3.10/dist-packages/torch/utils/checkpoint.py:429:
UserWarning: torch.utils.checkpoint: please pass in use_reentrant=True or
use_reentrant=False explicitly. The default value of use_reentrant will be
updated to be False in the future. To maintain current behavior, pass
use_reentrant=True. It is recommended that you use use_reentrant=False. Refer to
docs for more details on the differences between the two variants.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/transformers/generation/utils.py:1273:
UserWarning: Using the model-agnostic default `max_length` (=20) to control the
generation length. We recommend setting `max_new_tokens` to control the maximum
length of the generation.
  warnings.warn(
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
/usr/local/lib/python3.10/dist-packages/torch/utils/checkpoint.py:429:
UserWarning: torch.utils.checkpoint: please pass in use_reentrant=True or
use_reentrant=False explicitly. The default value of use_reentrant will be
updated to be False in the future. To maintain current behavior, pass
use_reentrant=True. It is recommended that you use use_reentrant=False. Refer to
docs for more details on the differences between the two variants.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/transformers/generation/utils.py:1273:
UserWarning: Using the model-agnostic default `max_length` (=20) to control the
generation length. We recommend setting `max_new_tokens` to control the maximum
length of the generation.
  warnings.warn(
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
/usr/local/lib/python3.10/dist-packages/torch/utils/checkpoint.py:429:
UserWarning: torch.utils.checkpoint: please pass in use_reentrant=True or
```

```
use_reentrant=False explicitly. The default value of use_reentrant will be
updated to be False in the future. To maintain current behavior, pass
use_reentrant=True. It is recommended that you use use_reentrant=False. Refer to
docs for more details on the differences between the two variants.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/transformers/generation/utils.py:1273:
UserWarning: Using the model-agnostic default `max_length` (=20) to control the
generation length. We recommend setting `max_new_tokens` to control the maximum
length of the generation.
  warnings.warn(
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
/usr/local/lib/python3.10/dist-packages/torch/utils/checkpoint.py:429:
UserWarning: torch.utils.checkpoint: please pass in use_reentrant=True or
use_reentrant=False explicitly. The default value of use_reentrant will be
updated to be False in the future. To maintain current behavior, pass
use_reentrant=True. It is recommended that you use use_reentrant=False. Refer to
docs for more details on the differences between the two variants.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/transformers/generation/utils.py:1273:
UserWarning: Using the model-agnostic default `max_length` (=20) to control the
generation length. We recommend setting `max_new_tokens` to control the maximum
length of the generation.
  warnings.warn(
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
/usr/local/lib/python3.10/dist-packages/torch/utils/checkpoint.py:429:
UserWarning: torch.utils.checkpoint: please pass in use_reentrant=True or
use_reentrant=False explicitly. The default value of use_reentrant will be
updated to be False in the future. To maintain current behavior, pass
use_reentrant=True. It is recommended that you use use_reentrant=False. Refer to
docs for more details on the differences between the two variants.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/transformers/generation/utils.py:1273:
UserWarning: Using the model-agnostic default `max_length` (=20) to control the
generation length. We recommend setting `max_new_tokens` to control the maximum
length of the generation.
  warnings.warn(
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
```

```
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
/usr/local/lib/python3.10/dist-packages/torch/utils/checkpoint.py:429:
UserWarning: torch.utils.checkpoint: please pass in use_reentrant=True or
use_reentrant=False explicitly. The default value of use_reentrant will be
updated to be False in the future. To maintain current behavior, pass
use_reentrant=True. It is recommended that you use use_reentrant=False. Refer to
docs for more details on the differences between the two variants.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/transformers/generation/utils.py:1273:
UserWarning: Using the model-agnostic default `max_length` (=20) to control the
generation length. We recommend setting `max_new_tokens` to control the maximum
length of the generation.
  warnings.warn(
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
/usr/local/lib/python3.10/dist-packages/torch/utils/checkpoint.py:429:
UserWarning: torch.utils.checkpoint: please pass in use_reentrant=True or
use_reentrant=False explicitly. The default value of use_reentrant will be
updated to be False in the future. To maintain current behavior, pass
use_reentrant=True. It is recommended that you use use_reentrant=False. Refer to
docs for more details on the differences between the two variants.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/transformers/generation/utils.py:1273:
UserWarning: Using the model-agnostic default `max_length` (=20) to control the
generation length. We recommend setting `max_new_tokens` to control the maximum
length of the generation.
  warnings.warn(
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
/usr/local/lib/python3.10/dist-packages/torch/utils/checkpoint.py:429:
UserWarning: torch.utils.checkpoint: please pass in use_reentrant=True or
use_reentrant=False explicitly. The default value of use_reentrant will be
updated to be False in the future. To maintain current behavior, pass
use_reentrant=True. It is recommended that you use use_reentrant=False. Refer to
docs for more details on the differences between the two variants.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/transformers/generation/utils.py:1273:
UserWarning: Using the model-agnostic default `max_length` (=20) to control the
generation length. We recommend setting `max_new_tokens` to control the maximum
length of the generation.
  warnings.warn(
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
```

```
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
/usr/local/lib/python3.10/dist-packages/torch/utils/checkpoint.py:429:
UserWarning: torch.utils.checkpoint: please pass in use_reentrant=True or
use_reentrant=False explicitly. The default value of use_reentrant will be
updated to be False in the future. To maintain current behavior, pass
use_reentrant=True. It is recommended that you use use_reentrant=False. Refer to
docs for more details on the differences between the two variants.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/transformers/generation/utils.py:1273:
UserWarning: Using the model-agnostic default `max_length` (=20) to control the
generation length. We recommend setting `max_new_tokens` to control the maximum
length of the generation.
  warnings.warn(
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
/usr/local/lib/python3.10/dist-packages/torch/utils/checkpoint.py:429:
UserWarning: torch.utils.checkpoint: please pass in use_reentrant=True or
use_reentrant=False explicitly. The default value of use_reentrant will be
updated to be False in the future. To maintain current behavior, pass
use_reentrant=True. It is recommended that you use use_reentrant=False. Refer to
docs for more details on the differences between the two variants.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/transformers/generation/utils.py:1273:
UserWarning: Using the model-agnostic default `max_length` (=20) to control the
generation length. We recommend setting `max_new_tokens` to control the maximum
length of the generation.
  warnings.warn(
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
/usr/local/lib/python3.10/dist-packages/torch/utils/checkpoint.py:429:
UserWarning: torch.utils.checkpoint: please pass in use_reentrant=True or
use_reentrant=False explicitly. The default value of use_reentrant will be
updated to be False in the future. To maintain current behavior, pass
use_reentrant=True. It is recommended that you use use_reentrant=False. Refer to
docs for more details on the differences between the two variants.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/transformers/generation/utils.py:1273:
UserWarning: Using the model-agnostic default `max_length` (=20) to control the
generation length. We recommend setting `max_new_tokens` to control the maximum
length of the generation.
  warnings.warn(
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
```

```
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
/usr/local/lib/python3.10/dist-packages/torch/utils/checkpoint.py:429:
UserWarning: torch.utils.checkpoint: please pass in use_reentrant=True or
use_reentrant=False explicitly. The default value of use_reentrant will be
updated to be False in the future. To maintain current behavior, pass
use_reentrant=True. It is recommended that you use use_reentrant=False. Refer to
docs for more details on the differences between the two variants.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/transformers/generation/utils.py:1273:
UserWarning: Using the model-agnostic default `max_length` (=20) to control the
generation length. We recommend setting `max_new_tokens` to control the maximum
length of the generation.
  warnings.warn(
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
/usr/local/lib/python3.10/dist-packages/torch/utils/checkpoint.py:429:
UserWarning: torch.utils.checkpoint: please pass in use_reentrant=True or
use_reentrant=False explicitly. The default value of use_reentrant will be
updated to be False in the future. To maintain current behavior, pass
use_reentrant=True. It is recommended that you use use_reentrant=False. Refer to
docs for more details on the differences between the two variants.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/transformers/generation/utils.py:1273:
UserWarning: Using the model-agnostic default `max_length` (=20) to control the
generation length. We recommend setting `max_new_tokens` to control the maximum
length of the generation.
  warnings.warn(
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
/usr/local/lib/python3.10/dist-packages/torch/utils/checkpoint.py:429:
UserWarning: torch.utils.checkpoint: please pass in use_reentrant=True or
use_reentrant=False explicitly. The default value of use_reentrant will be
updated to be False in the future. To maintain current behavior, pass
use_reentrant=True. It is recommended that you use use_reentrant=False. Refer to
docs for more details on the differences between the two variants.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/transformers/generation/utils.py:1273:
UserWarning: Using the model-agnostic default `max_length` (=20) to control the
generation length. We recommend setting `max_new_tokens` to control the maximum
length of the generation.
  warnings.warn(
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
/usr/local/lib/python3.10/dist-packages/torch/utils/checkpoint.py:429:
UserWarning: torch.utils.checkpoint: please pass in use_reentrant=True or
```

```
use_reentrant=False explicitly. The default value of use_reentrant will be
updated to be False in the future. To maintain current behavior, pass
use_reentrant=True. It is recommended that you use use_reentrant=False. Refer to
docs for more details on the differences between the two variants.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/transformers/generation/utils.py:1273:
UserWarning: Using the model-agnostic default `max_length` (=20) to control the
generation length. We recommend setting `max_new_tokens` to control the maximum
length of the generation.
  warnings.warn(
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
/usr/local/lib/python3.10/dist-packages/torch/utils/checkpoint.py:429:
UserWarning: torch.utils.checkpoint: please pass in use_reentrant=True or
use_reentrant=False explicitly. The default value of use_reentrant will be
updated to be False in the future. To maintain current behavior, pass
use_reentrant=True. It is recommended that you use use_reentrant=False. Refer to
docs for more details on the differences between the two variants.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/transformers/generation/utils.py:1273:
UserWarning: Using the model-agnostic default `max_length` (=20) to control the
generation length. We recommend setting `max_new_tokens` to control the maximum
length of the generation.
  warnings.warn(
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
/usr/local/lib/python3.10/dist-packages/torch/utils/checkpoint.py:429:
UserWarning: torch.utils.checkpoint: please pass in use_reentrant=True or
use_reentrant=False explicitly. The default value of use_reentrant will be
updated to be False in the future. To maintain current behavior, pass
use_reentrant=True. It is recommended that you use use_reentrant=False. Refer to
docs for more details on the differences between the two variants.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/transformers/generation/utils.py:1273:
UserWarning: Using the model-agnostic default `max_length` (=20) to control the
generation length. We recommend setting `max_new_tokens` to control the maximum
length of the generation.
  warnings.warn(
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
/usr/local/lib/python3.10/dist-packages/torch/utils/checkpoint.py:429:
UserWarning: torch.utils.checkpoint: please pass in use_reentrant=True or
use_reentrant=False explicitly. The default value of use_reentrant will be
updated to be False in the future. To maintain current behavior, pass
use_reentrant=True. It is recommended that you use use_reentrant=False. Refer to
docs for more details on the differences between the two variants.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/transformers/generation/utils.py:1273:
```

```
UserWarning: Using the model-agnostic default `max_length` (=20) to control the
generation length. We recommend setting `max_new_tokens` to control the maximum
length of the generation.
  warnings.warn(
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
/usr/local/lib/python3.10/dist-packages/torch/utils/checkpoint.py:429:
UserWarning: torch.utils.checkpoint: please pass in use_reentrant=True or
use_reentrant=False explicitly. The default value of use_reentrant will be
updated to be False in the future. To maintain current behavior, pass
use_reentrant=True. It is recommended that you use use_reentrant=False. Refer to
docs for more details on the differences between the two variants.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/transformers/generation/utils.py:1273:
UserWarning: Using the model-agnostic default `max_length` (=20) to control the
generation length. We recommend setting `max_new_tokens` to control the maximum
length of the generation.
  warnings.warn(
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
/usr/local/lib/python3.10/dist-packages/torch/utils/checkpoint.py:429:
UserWarning: torch.utils.checkpoint: please pass in use_reentrant=True or
use_reentrant=False explicitly. The default value of use_reentrant will be
updated to be False in the future. To maintain current behavior, pass
use_reentrant=True. It is recommended that you use use_reentrant=False. Refer to
docs for more details on the differences between the two variants.
  warnings.warn(
There were missing keys in the checkpoint model loaded:
['encoder.embed_tokens.weight', 'decoder.embed_tokens.weight',
'lm_head.weight'].
/usr/local/lib/python3.10/dist-packages/transformers/generation/utils.py:1273:
UserWarning: Using the model-agnostic default `max_length` (=20) to control the
generation length. We recommend setting `max_new_tokens` to control the maximum
length of the generation.
  warnings.warn(
```

```
<IPython.core.display.HTML object>
```

```
The best model was saved at step 250.
```

```
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
Warning: Empty candidate sentence detected; setting raw BERTscores to 0.
```

```
<IPython.core.display.HTML object>
```

```
<IPython.core.display.HTML object>
```

```
<IPython.core.display.HTML object>
```

```
<IPython.core.display.HTML object>
```

[31]:

```
model_folder = base_folder / "models" / "nlp_spring_2023/kde4/opus-mt-en-fr"

checkpoint = str(model_folder / "checkpoint-{}".format(best_model))
test_data_flattened = test_subset["test"].map(lambda example: {'en':␣
 ↪example['translation']['en'], 'fr': example['translation']['fr']})



task_evaluator = evaluator("translation")
gen_kwargs = {"length_penalty": 0.8, "num_beams": 8, "max_length": 128}
eval_results = task_evaluator.compute(
    model_or_pipeline=checkpoint,
    tokenizer=checkpoint,
    data=test_data_flattened,
    input_column='en',
    label_column='fr',
    generation_kwargs=gen_kwargs,
    device=0,
)

eval_results
```

```
/usr/local/lib/python3.10/dist-packages/transformers/pipelines/__init__.py:1049:
UserWarning: "translation" task was used, instead of "translation_XX_to_YY",
defaulting to "translation_en_to_de"
  warnings.warn(
Your input_length: 144 is bigger than 0.9 * max_length: 128. You might consider
increasing your max_length manually, e.g. translator('…', max_length=400)
```

[31]:

```
{'bleu': 0.06705230814129132,
 'precisions': [0.24748883928571427,
  0.12062256809338522,
  0.08123145400593472,
  0.052742616033755275],
 'brevity_penalty': 0.6305161446131688,
 'length_ratio': 0.6843612755394309,
 'translation_length': 3584,
 'reference_length': 5237,
 'total_time_in_seconds': 135.197146945,
 'samples_per_second': 3.698302895425794,
 'latency_in_seconds': 0.27039429388999997}
```

0.4.2 Experiment 2 - Helsinki - English to French, Batch Size = 32

[18]:

```
from transformers import Seq2SeqTrainingArguments

# Define the directory where model checkpoints will be saved
model_folder = base_folder / "models" / "nlp_spring_2023/kde4/opus-mt-en-fr"

# Create the directory if it doesn't exist
model_folder.mkdir(exist_ok=True, parents=True)

checkpoint = 'Helsinki-NLP/opus-mt-en-fr'
generation_config = GenerationConfig.from_pretrained(checkpoint)
tokenizer = AutoTokenizer.from_pretrained(checkpoint)

# Configure training parameters
training_args = Seq2SeqTrainingArguments(
    # Training-specific configurations
    num_train_epochs=1,  # Total number of training epochs
    weight_decay=0.01,  # Apply L2 regularization to prevent overfitting
    learning_rate=5e-5,  # Step size for the optimizer during training
    optim="adamw_torch",  # Optimizer,
    warmup_steps=10,
    predict_with_generate=True,
    #generation_config=generation_config,
    # memory and speed related arguments
    # Number of samples per training batch for each device
    per_device_train_batch_size=32,
    per_device_eval_batch_size=32,  # Number of samples per eval batch for each␣
 ↪device

    gradient_checkpointing=True,  # memory
    # fp16 = True, # Speed
    # bf16=True,
    # tf32=True, # speed
    # evaluation settings
    output_dir=str(model_folder),  # Directory to save model checkpoints
    evaluation_strategy="steps",  # Evaluate model at specified step intervals
    eval_steps=10,  # Perform evaluation every 10 training steps
    # Checkpoint settings
    save_strategy="steps",  # Save model checkpoint at specified step intervals
    save_steps=10,  # Save a model checkpoint every 10 training steps
    load_best_model_at_end=True,  # Reload the best model at the end of training
    save_total_limit=2,  # Retain only the best and the most recent model␣
 ↪checkpoints
    # metric_for_best_model=,
    # greater_is_better=,
    # Experiment logging configurations (commented out in this example)
    logging_strategy="steps",
    logging_steps=10,
    report_to="wandb",  # Log metrics and results to Weights & Biases platform
    # Experiment name for Weights & Biases
    run_name="translation-exp2",
)
```

```
/usr/local/lib/python3.10/dist-
packages/transformers/models/marian/tokenization_marian.py:197: UserWarning:
Recommended: pip install sacremoses.
  warnings.warn("Recommended: pip install sacremoses.")
```

[19]:

```
train_val_subset, test_subset = split_dataset()
```

[20]:

```
training_args_dict = training_args.to_dict() # Convert TrainingArguments to␣
  ↪dictionary
#training_args_dict['run_name'] = f'{checkpoint}' # Update the run_name
new_training_args = Seq2SeqTrainingArguments(**training_args_dict)
```

```
/usr/local/lib/python3.10/dist-packages/transformers/training_args.py:1697:
FutureWarning: `--push_to_hub_token` is deprecated and will be removed in
version 5 of   Transformers. Use `--hub_token` instead.
  warnings.warn(
```

[21]:

```
best_model = tokenize_train_evaluate_log(training_args= new_training_args,␣
  ↪checkpoint=checkpoint, base_folder=base_folder,␣
  ↪train_val_subset=train_val_subset,
                             compute_metrics=compute_metrics)
```

```
env: WANDB_PROJECT=nlp_course_fall_2023-HW6-PartD

<IPython.core.display.HTML object>

<IPython.core.display.HTML object>

<IPython.core.display.HTML object>

<IPython.core.display.HTML object>

<IPython.core.display.HTML object>

<IPython.core.display.HTML object>

<IPython.core.display.HTML object>

/usr/local/lib/python3.10/dist-packages/torch/utils/checkpoint.py:429:
UserWarning: torch.utils.checkpoint: please pass in use_reentrant=True or
use_reentrant=False explicitly. The default value of use_reentrant will be
updated to be False in the future. To maintain current behavior, pass
use_reentrant=True. It is recommended that you use use_reentrant=False. Refer to
docs for more details on the differences between the two variants.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/torch/utils/checkpoint.py:429:
UserWarning: torch.utils.checkpoint: please pass in use_reentrant=True or
use_reentrant=False explicitly. The default value of use_reentrant will be
updated to be False in the future. To maintain current behavior, pass
use_reentrant=True. It is recommended that you use use_reentrant=False. Refer to
docs for more details on the differences between the two variants.
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/torch/utils/checkpoint.py:429:
UserWarning: torch.utils.checkpoint: please pass in use_reentrant=True or
use_reentrant=False explicitly. The default value of use_reentrant will be
updated to be False in the future. To maintain current behavior, pass
use_reentrant=True. It is recommended that you use use_reentrant=False. Refer to
docs for more details on the differences between the two variants.
  warnings.warn(
There were missing keys in the checkpoint model loaded:
['model.encoder.embed_tokens.weight', 'model.encoder.embed_positions.weight',
'model.decoder.embed_tokens.weight', 'model.decoder.embed_positions.weight',
'lm_head.weight'].
```

```
<IPython.core.display.HTML object>
```

```
The best model was saved at step 30.
```

```
<IPython.core.display.HTML object>
```

```
<IPython.core.display.HTML object>
```

```
<IPython.core.display.HTML object>
```

```
<IPython.core.display.HTML object>
```

[22]:

```
model_folder = base_folder / "models" / "nlp_spring_2023/kde4/opus-mt-en-fr"

checkpoint = str(model_folder / "checkpoint-30")
test_data_flattened = test_subset["test"].map(lambda example: {'en': example['translation']['en'], 'fr': example['translation']['fr']})




task_evaluator = evaluator("translation")
gen_kwargs = {"length_penalty": 0.8, "num_beams": 8, "max_length": 128}
eval_results = task_evaluator.compute(
    model_or_pipeline=checkpoint,
    tokenizer=checkpoint,
    data=test_data_flattened,
    input_column='en',
    label_column='fr',
    generation_kwargs=gen_kwargs,
    device=0,
)
```

```
/usr/local/lib/python3.10/dist-
packages/transformers/models/marian/tokenization_marian.py:197: UserWarning:
Recommended: pip install sacremoses.
  warnings.warn("Recommended: pip install sacremoses.")
```

```
Downloading builder script:   0%|          | 0.00/5.94k [00:00<?, ?B/s]
```

```
Downloading extra modules:   0%|          | 0.00/1.55k [00:00<?, ?B/s]
```

```
Downloading extra modules:   0%|          | 0.00/3.34k [00:00<?, ?B/s]
```

```
Your input_length: 130 is bigger than 0.9 * max_length: 128. You might consider
increasing your max_length manually, e.g. translator('…', max_length=400)
```

[23]:

```
eval_results
```

[23]:

```
{'bleu': 0.42319476783921356,
 'precisions': [0.706183368869936,
  0.5322195704057279,
  0.4186292670018523,
  0.32503660322108346],
 'brevity_penalty': 0.8899133949248703,
 'length_ratio': 0.8955508879129273,
 'translation_length': 4690,
 'reference_length': 5237,
 'total_time_in_seconds': 108.61369759799982,
 'samples_per_second': 4.603470934675258,
 'latency_in_seconds': 0.21722739519599962}
```

Here I ran 2 experiment on Machine Translation task

1. I changed the model to T5-Base with 4 epochs which translates English to German and is smaller model compared to Helsinki NLP and hence gives terrible Bleu Score of 0.06705
2. I increased batch size from 16 to 32 with the same Helsinki NLP model and didn't see drastic change in the model performance over test or evaluation dataset with Bleu score of 0.4232

[ ]:

```

```