

Modified from DeepLearning.AI Hugging Face course

<https://learn.deeplearning.ai/courses/open-source-models-hugging-face/lesson/3/natural-language-processing-nlp>

```
!pip install transformers
```

```
!pip install torch
```

```
Requirement already satisfied: transformers in /usr/local/lib/python3.10/dist-packages (4.38.2)
Requirement already satisfied: filelock in /usr/local/lib/python3.10/dist-packages (from transformers)
Requirement already satisfied: huggingface-hub<1.0,>=0.19.3 in /usr/local/lib/python3.10/dist-packages (from transformers)
Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.10/dist-packages (from transformers)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist-packages (from transformers)
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.10/dist-packages (from transformers)
Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/python3.10/dist-packages (from transformers)
Requirement already satisfied: requests in /usr/local/lib/python3.10/dist-packages (from transformers)
Requirement already satisfied: tokenizers<0.19,>=0.14 in /usr/local/lib/python3.10/dist-packages (from transformers)
Requirement already satisfied: safetensors>=0.4.1 in /usr/local/lib/python3.10/dist-packages (from transformers)
Requirement already satisfied: tqdm>=4.27 in /usr/local/lib/python3.10/dist-packages (from transformers)
Requirement already satisfied: fsspec>=2023.5.0 in /usr/local/lib/python3.10/dist-packages (from transformers)
Requirement already satisfied: typing-extensions>=3.7.4.3 in /usr/local/lib/python3.10/dist-packages (from transformers)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from transformers)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from transformers)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from transformers)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from transformers)
```

```
# suppress warnings
```

```
from transformers.utils import logging
logging.set_verbosity_error()
```

```
from transformers import pipeline
import torch
```

```
# bfloat16 data type will compress the model without degradation
```

```
translator = pipeline(task="translation",
                      model="facebook/nllb-200-distilled-600M",
                      torch_dtype=torch.bfloat16)
```

```
/usr/local/lib/python3.10/dist-packages/huggingface_hub/utils/_token.py:88: 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://huggingface.co/settings/tokens).
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 models or datasets
warnings.warn(
```

```
config.json: 100% 846/846 [00:00<00:00, 11.4kB/s]
```

```
pytorch_model.bin: 100% 2.46G/2.46G [00:36<00:00, 73.0MB/s]
```

```
/usr/local/lib/python3.10/dist-packages/torch/_utils.py:831: UserWarning: TypedStorage is deprecated
  return self.fget.__get__(instance, owner)()
```

```
generation_config.json: 100% 189/189 [00:00<00:00, 8.58kB/s]
```

```
tokenizer_config.json: 100% 564/564 [00:00<00:00, 23.8kB/s]
```

```
sentencepiece.bpe.model: 100% 4.85M/4.85M [00:05<00:00, 838kB/s]
```

```
tokenizer.json: 100% 17.3M/17.3M [00:02<00:00, 8.43MB/s]
```

```
special_tokens_map.json: 100% 3.55k/3.55k [00:00<00:00, 175kB/s]
```

```
text = """\
My puppy is adorable, \
Your kitten is cute.
Her panda is friendly.
His llama is thoughtful. \
We all have nice pets!"""
```

```
text_translated = translator(text,
                             src_lang="eng_Latn",
                             tgt_lang="fra_Latn")
```

To see the available languages, go to:

<https://github.com/facebookresearch/flores/blob/main/flores200/README.md#languages-in-flores-200>

```
text_translated
```

```
[{'translation_text': 'Mon chiot est adorable, ton chaton est mignon, son panda est ami, sa lamme
est attentive, nous avons tous de beaux animaux de compagnie.'}]
```

```
# free up some memory
```

```
import gc
```

```
del translator
```

```
gc.collect()
```

```
103
```

```
# build the summarization pipeline
```

```
summarizer = pipeline(task="summarization",
                      model="facebook/bart-large-cnn",
                      torch_dtype=torch.bfloat16)
```

config.json: 100%	1.58k/1.58k [00:00<00:00, 30.9kB/s]
model.safetensors: 100%	1.63G/1.63G [00:24<00:00, 59.5MB/s]
generation_config.json: 100%	363/363 [00:00<00:00, 5.24kB/s]
vocab.json: 100%	899k/899k [00:00<00:00, 1.13MB/s]
merges.txt: 100%	456k/456k [00:00<00:00, 766kB/s]
tokenizer.json: 100%	1.36M/1.36M [00:00<00:00, 1.37MB/s]

```
text = """Paris is the capital and most populous city of France, with
an estimated population of 2,175,601 residents as of 2018,
in an area of more than 105 square kilometres (41 square
miles). The City of Paris is the centre and seat of
government of the region and province of Île-de-France, or
Paris Region, which has an estimated population of
12,174,880, or about 18 percent of the population of France
as of 2017."""
```

```
summary = summarizer(text,  
                      min_length=10,  
                      max_length=100)
```

```
summary
```

```
[{'summary_text': 'Paris is the capital and most populous city of France, with an estimated  
population of 2,175,601 residents as of 2018. The City of Paris is the centre and seat of the  
government of the region and province of Île-de-France.'}]
```