# Project Report- Building a Multilingual Speech Recognition Model for RAG Without Training

## 1. Introduction

Speech is a core human capability for articulating and sharing thoughts, serving as a natural and efficient medium for communication. Speech recognition technology translates spoken words into text, enabling users to interact with electronic devices through voice commands. Similarly, speech translation technology facilitates communication across different languages, bridging language barriers. Advancements in technology have led to the development of multilingual speech recognition models powered by artificial intelligence. These models enhance user interfaces with additional functionalities and find applications across various fields such as science and education. When integrated with Retrieval-Augmented Generation (RAG) systems, these models can create voice-activated solutions that offer a more dynamic and interactive user experience.

### 1.1 Objective

A multilingual speech recognition model is designed to handle audio input by identifying the source language and translating it into the desired target language. This model leverages deep learning techniques to achieve high accuracy. The development of such a model involves several key components: recognizing speech, verifying the source language, converting speech to text, translating the text, and then converting it into the target language. By integrating this model with a pre-trained Retrieval-Augmented Generation (RAG) model from Hugging Face's Transformers library, the final system can generate responses based on the retrieved documents, enhancing its ability to provide relevant and contextually accurate replies.

### 1.2 Background

**Multilingual Speech Recognition**
Multilingual speech recognition, especially when enhanced through supervised learning, has achieved impressive results and garnered substantial interest due to its capability to transcribe multiple languages. AI-driven models from leading technology firms like Google and Microsoft are at the forefront of this technology, enabling effective language translation.

**Retrieval-Augmented Generation (RAG)**
RAG is an advanced AI framework that combines the strengths of traditional information retrieval systems with the capabilities of generative large language models. When a query is input, the RAG retriever searches through a vast database to select the most pertinent documents. This retrieved information is then used by the model to generate coherent and contextually appropriate responses.

**Integration of Speech Recognition and RAG**
Combining multilingual speech recognition with RAG creates a powerful system capable of translating, transcribing, and retrieving information in multiple languages. Pretrained models for both speech recognition and translation, such as Whisper, known for its high transcription accuracy, can be integrated. This fusion results in a sophisticated multilingual speech recognition system that leverages the strengths of both technologies.

## 2. Methodology

### 2.1. Model Selection
Whisper is an advanced automated speech recognition (ASR) system developed by OpenAI, trained on 680,000 hours of multilingual supervised data sourced from the web. This extensive and varied training dataset enhances Whisper's robustness and accuracy in speech processing. It is capable of transcribing speech in multiple languages and translating these transcriptions into English.

### 2.2. Audio Processing
The Whisper model utilizes an encoder-decoder Transformer architecture. Audio input is processed by the Whisper model, where it is first divided into 30-second segments. Each segment is converted into a log-Mel spectrogram, which is then fed into the encoder. The decoder is trained to generate corresponding text captions, incorporating special tokens to guide the model in performing tasks such as language identification, generating phrase-level timestamps, multilingual transcription, and translating speech to English.
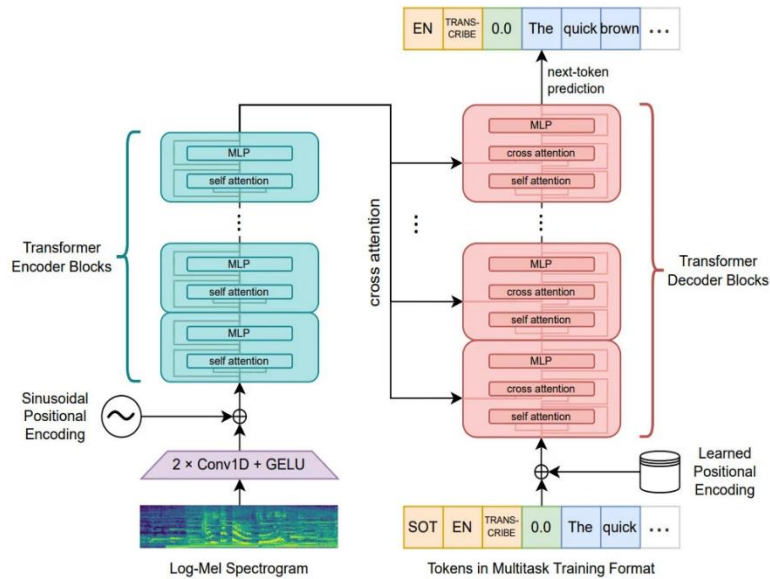
Fig 1. Speech Processing in Whisper

### 2.1. Translation Tool
For translation, the Google Translator tool from the 'deep-translator' library is employed. This tool is versatile and provides seamless translation between a variety of languages using multiple translation engines. Its ease of use and support for well-known universal translators make it a valuable resource for language translation tasks.

### 2.2. Document Retrieval
A collection of pre-defined dummy documents, relevant to multilingual speech recognition queries, has been established. Due to resource constraints and limited environmental support, these documents are managed manually rather than through advanced retrieval libraries.

### 2.3. Response Generation
The RagSequenceForGeneration model from Hugging Face's Transformers library is

utilized for response generation. This model executes RAG-sequence-specific marginalization during the forward pass and includes two primary components: a question encoder and a generator. The question encoder is an autoencoding model called DPRQuestionEncoder, which is a subclass of PreTrainedModel.

- config (RagConfig): This class manages the model's configuration parameters. It initializes with a configuration file but does not load the model weights; use the from_pretrained() method to load the weights.
- question_encoder (PreTrainedModel): An encoder model compatible with the FAISS index used by the retriever.
- generator (PreTrainedModel): A sequence-to-sequence model serving as the generator within the RAG framework.
- retriever (RagRetriever): A class that integrates a FAISS index to query and obtain contextual documents relevant to the current inputs.

## 2.1. Integration with User Interface

Gradio, an opensource python package, is used to build a demo web application to create user friendly interaction with the model.
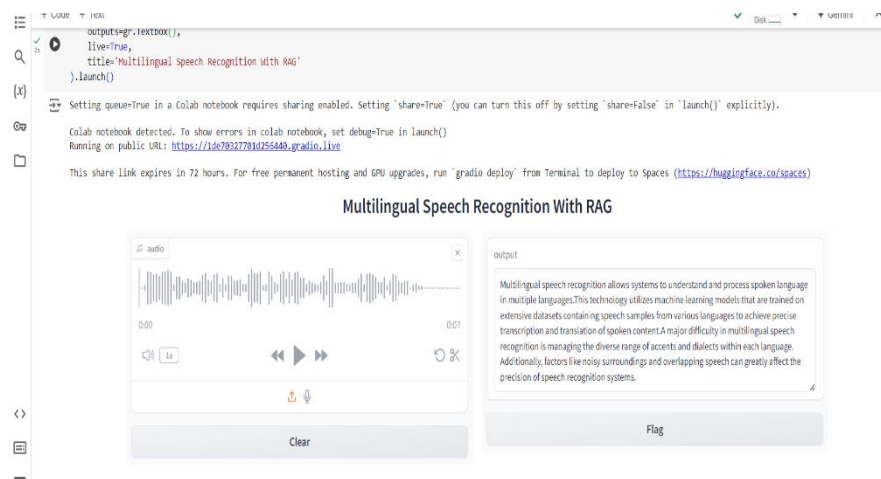


Fig 2. Snapshot of Interface

## 3. Evaluation

### 3.1. Speech Recognition Accuracy

Word Error Rate (WER) is a common metric of the performance of a speech recognition. The WER was 2% for the model.

### 3.2. Translation Quality

BLUE (BiLingual Evaluation Understudy) is a metric for automatically evaluating machine-translated text. It is a number between zero and one that measures the similarity of the machine-translated text. The BLEU score for the translation system was calculated to be 1.00. This indicates the generated translations perfectly match the translations in terms of n-gram overlap.

### Conclusion

In conclusion, the project successfully developed a multilingual speech recognition and translation system with good accuracy in both transcription and translation tasks. For future

works, we can focus on expanding the datasets and adding relevant documents to further validate the model's performance.

## 4. References

- **https://deep-translator.readthedocs.io/en/latest/**
- **https://huggingface.co/docs/transformers/en/model_doc/rag**
- **https://www.gradio.app/docs/gradio/interface**

# ananya-tensorgo

August 6, 2024

```
[1]: !pip install git+https://github.com/openai/whisper.git -q
```

```
Installing build dependencies … done
Getting requirements to build wheel … done
Preparing metadata (pyproject.toml) … done
                          1.1/1.1 MB
32.4 MB/s eta 0:00:00
Building wheel for openai-whisper (pyproject.toml) … done
```

```
[3]: !pip install deep-translator
```

```
Collecting deep-translator
  Downloading deep_translator-1.11.4-py3-none-any.whl.metadata (30 kB)
Requirement already satisfied: beautifulsoup4<5.0.0,>=4.9.1 in
/usr/local/lib/python3.10/dist-packages (from deep-translator) (4.12.3)
Requirement already satisfied: requests<3.0.0,>=2.23.0 in
/usr/local/lib/python3.10/dist-packages (from deep-translator) (2.31.0)
Requirement already satisfied: soupsieve>1.2 in /usr/local/lib/python3.10/dist-
packages (from beautifulsoup4<5.0.0,>=4.9.1->deep-translator) (2.5)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.10/dist-packages (from requests<3.0.0,>=2.23.0->deep-
translator) (3.3.2)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-
packages (from requests<3.0.0,>=2.23.0->deep-translator) (3.7)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.10/dist-packages (from requests<3.0.0,>=2.23.0->deep-
translator) (2.0.7)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.10/dist-packages (from requests<3.0.0,>=2.23.0->deep-
translator) (2024.7.4)
Downloading deep_translator-1.11.4-py3-none-any.whl (42 kB)
                          42.3/42.3 kB
2.5 MB/s eta 0:00:00
Installing collected packages: deep-translator
Successfully installed deep-translator-1.11.4
```

```
[ ]: !pip install --upgrade gradio
```

```
Collecting gradio
  Downloading gradio-4.40.0-py3-none-any.whl.metadata (15 kB)
Collecting aiofiles<24.0,>=22.0 (from gradio)
  Downloading aiofiles-23.2.1-py3-none-any.whl.metadata (9.7 kB)
Requirement already satisfied: anyio<5.0,>=3.0 in
/usr/local/lib/python3.10/dist-packages (from gradio) (3.7.1)
Collecting fastapi (from gradio)
  Downloading fastapi-0.112.0-py3-none-any.whl.metadata (27 kB)
Collecting ffmpy (from gradio)
  Downloading ffmpy-0.4.0-py3-none-any.whl.metadata (2.9 kB)
Collecting gradio-client==1.2.0 (from gradio)
  Downloading gradio_client-1.2.0-py3-none-any.whl.metadata (7.1 kB)
Collecting httpx>=0.24.1 (from gradio)
  Downloading httpx-0.27.0-py3-none-any.whl.metadata (7.2 kB)
Requirement already satisfied: huggingface-hub>=0.19.3 in
/usr/local/lib/python3.10/dist-packages (from gradio) (0.23.5)
Requirement already satisfied: importlib-resources<7.0,>=1.3 in
/usr/local/lib/python3.10/dist-packages (from gradio) (6.4.0)
Requirement already satisfied: jinja2<4.0 in /usr/local/lib/python3.10/dist-
packages (from gradio) (3.1.4)
Requirement already satisfied: markupsafe~=2.0 in
/usr/local/lib/python3.10/dist-packages (from gradio) (2.1.5)
Requirement already satisfied: matplotlib~=3.0 in
/usr/local/lib/python3.10/dist-packages (from gradio) (3.7.1)
Requirement already satisfied: numpy<3.0,>=1.0 in
/usr/local/lib/python3.10/dist-packages (from gradio) (1.26.4)
Collecting orjson~=3.0 (from gradio)
  Downloading orjson-3.10.6-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_
64.whl.metadata (50 kB)
                         50.4/50.4 kB
2.3 MB/s eta 0:00:00
Requirement already satisfied: packaging in
/usr/local/lib/python3.10/dist-packages (from gradio) (24.1)
Requirement already satisfied: pandas<3.0,>=1.0 in
/usr/local/lib/python3.10/dist-packages (from gradio) (2.1.4)
Requirement already satisfied: pillow<11.0,>=8.0 in
/usr/local/lib/python3.10/dist-packages (from gradio) (9.4.0)
Requirement already satisfied: pydantic>=2.0 in /usr/local/lib/python3.10/dist-
packages (from gradio) (2.8.2)
Collecting pydub (from gradio)
  Downloading pydub-0.25.1-py2.py3-none-any.whl.metadata (1.4 kB)
Collecting python-multipart>=0.0.9 (from gradio)
  Downloading python_multipart-0.0.9-py3-none-any.whl.metadata (2.5 kB)
Requirement already satisfied: pyyaml<7.0,>=5.0 in
/usr/local/lib/python3.10/dist-packages (from gradio) (6.0.1)
Collecting ruff>=0.2.2 (from gradio)
  Downloading ruff-0.5.6-py3-none-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (24 kB)
```

```
Collecting semantic-version~=2.0 (from gradio)
  Downloading semantic_version-2.10.0-py2.py3-none-any.whl.metadata (9.7 kB)
Collecting tomlkit==0.12.0 (from gradio)
  Downloading tomlkit-0.12.0-py3-none-any.whl.metadata (2.7 kB)
Requirement already satisfied: typer<1.0,>=0.12 in
/usr/local/lib/python3.10/dist-packages (from gradio) (0.12.3)
Requirement already satisfied: typing-extensions~=4.0 in
/usr/local/lib/python3.10/dist-packages (from gradio) (4.12.2)
Requirement already satisfied: urllib3~=2.0 in /usr/local/lib/python3.10/dist-
packages (from gradio) (2.0.7)
Collecting uvicorn>=0.14.0 (from gradio)
  Downloading uvicorn-0.30.5-py3-none-any.whl.metadata (6.6 kB)
Requirement already satisfied: fsspec in /usr/local/lib/python3.10/dist-packages
(from gradio-client==1.2.0->gradio) (2024.6.1)
Collecting websockets<13.0,>=10.0 (from gradio-client==1.2.0->gradio)
  Downloading websockets-12.0-cp310-cp310-manylinux_2_5_x86_64.manylinux1_x86_64
.manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (6.6 kB)
Requirement already satisfied: idna>=2.8 in /usr/local/lib/python3.10/dist-
packages (from anyio<5.0,>=3.0->gradio) (3.7)
Requirement already satisfied: sniffio>=1.1 in /usr/local/lib/python3.10/dist-
packages (from anyio<5.0,>=3.0->gradio) (1.3.1)
Requirement already satisfied: exceptiongroup in /usr/local/lib/python3.10/dist-
packages (from anyio<5.0,>=3.0->gradio) (1.2.2)
Requirement already satisfied: certifi in /usr/local/lib/python3.10/dist-
packages (from httpx>=0.24.1->gradio) (2024.7.4)
Collecting httpcore==1.* (from httpx>=0.24.1->gradio)
  Downloading httpcore-1.0.5-py3-none-any.whl.metadata (20 kB)
Collecting h11<0.15,>=0.13 (from httpcore==1.*->httpx>=0.24.1->gradio)
  Downloading h11-0.14.0-py3-none-any.whl.metadata (8.2 kB)
Requirement already satisfied: filelock in /usr/local/lib/python3.10/dist-
packages (from huggingface-hub>=0.19.3->gradio) (3.15.4)
Requirement already satisfied: requests in /usr/local/lib/python3.10/dist-
packages (from huggingface-hub>=0.19.3->gradio) (2.31.0)
Requirement already satisfied: tqdm>=4.42.1 in /usr/local/lib/python3.10/dist-
packages (from huggingface-hub>=0.19.3->gradio) (4.66.4)
Requirement already satisfied: contourpy>=1.0.1 in
/usr/local/lib/python3.10/dist-packages (from matplotlib~=3.0->gradio) (1.2.1)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.10/dist-
packages (from matplotlib~=3.0->gradio) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in
/usr/local/lib/python3.10/dist-packages (from matplotlib~=3.0->gradio) (4.53.1)
Requirement already satisfied: kiwisolver>=1.0.1 in
/usr/local/lib/python3.10/dist-packages (from matplotlib~=3.0->gradio) (1.4.5)
Requirement already satisfied: pyparsing>=2.3.1 in
/usr/local/lib/python3.10/dist-packages (from matplotlib~=3.0->gradio) (3.1.2)
Requirement already satisfied: python-dateutil>=2.7 in
/usr/local/lib/python3.10/dist-packages (from matplotlib~=3.0->gradio) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-
```

packages (from pandas<3.0,>=1.0->gradio) (2024.1)
Requirement already satisfied: tzdata>=2022.1 in /usr/local/lib/python3.10/dist-
packages (from pandas<3.0,>=1.0->gradio) (2024.1)
Requirement already satisfied: annotated-types>=0.4.0 in
/usr/local/lib/python3.10/dist-packages (from pydantic>=2.0->gradio) (0.7.0)
Requirement already satisfied: pydantic-core==2.20.1 in
/usr/local/lib/python3.10/dist-packages (from pydantic>=2.0->gradio) (2.20.1)
Requirement already satisfied: click>=8.0.0 in /usr/local/lib/python3.10/dist-
packages (from typer<1.0,>=0.12->gradio) (8.1.7)
Requirement already satisfied: shellingham>=1.3.0 in
/usr/local/lib/python3.10/dist-packages (from typer<1.0,>=0.12->gradio) (1.5.4)
Requirement already satisfied: rich>=10.11.0 in /usr/local/lib/python3.10/dist-
packages (from typer<1.0,>=0.12->gradio) (13.7.1)
Collecting starlette<0.38.0,>=0.37.2 (from fastapi->gradio)
  Downloading starlette-0.37.2-py3-none-any.whl.metadata (5.9 kB)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-
packages (from python-dateutil>=2.7->matplotlib~=3.0->gradio) (1.16.0)
Requirement already satisfied: markdown-it-py>=2.2.0 in
/usr/local/lib/python3.10/dist-packages (from
rich>=10.11.0->typer<1.0,>=0.12->gradio) (3.0.0)
Requirement already satisfied: pygments<3.0.0,>=2.13.0 in
/usr/local/lib/python3.10/dist-packages (from
rich>=10.11.0->typer<1.0,>=0.12->gradio) (2.16.1)
Requirement already satisfied: charset-normalizer<4,>=2 in
/usr/local/lib/python3.10/dist-packages (from requests->huggingface-
hub>=0.19.3->gradio) (3.3.2)
Requirement already satisfied: mdurl~=0.1 in /usr/local/lib/python3.10/dist-
packages (from markdown-it-py>=2.2.0->rich>=10.11.0->typer<1.0,>=0.12->gradio)
(0.1.2)
Downloading gradio-4.40.0-py3-none-any.whl (12.5 MB)
                         12.5/12.5 MB
28.3 MB/s eta 0:00:00
Downloading gradio_client-1.2.0-py3-none-any.whl (318 kB)
                         318.6/318.6 kB
16.5 MB/s eta 0:00:00
Downloading tomlkit-0.12.0-py3-none-any.whl (37 kB)
Downloading aiofiles-23.2.1-py3-none-any.whl (15 kB)
Downloading httpx-0.27.0-py3-none-any.whl (75 kB)
                         75.6/75.6 kB
5.3 MB/s eta 0:00:00
Downloading httpcore-1.0.5-py3-none-any.whl (77 kB)
                         77.9/77.9 kB
2.6 MB/s eta 0:00:00
Downloading
orjson-3.10.6-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (141
kB)
                         141.1/141.1 kB
8.5 MB/s eta 0:00:00

```
Downloading python_multipart-0.0.9-py3-none-any.whl (22 kB)
Downloading ruff-0.5.6-py3-none-manylinux_2_17_x86_64.manylinux2014_x86_64.whl
(10.2 MB)
                              10.2/10.2 MB
29.2 MB/s eta 0:00:00
Downloading semantic_version-2.10.0-py2.py3-none-any.whl (15 kB)
Downloading uvicorn-0.30.5-py3-none-any.whl (62 kB)
                              62.8/62.8 kB
3.9 MB/s eta 0:00:00
Downloading fastapi-0.112.0-py3-none-any.whl (93 kB)
                              93.1/93.1 kB
3.0 MB/s eta 0:00:00
Downloading ffmpy-0.4.0-py3-none-any.whl (5.8 kB)
Downloading pydub-0.25.1-py2.py3-none-any.whl (32 kB)
Downloading h11-0.14.0-py3-none-any.whl (58 kB)
                              58.3/58.3 kB
3.7 MB/s eta 0:00:00
Downloading starlette-0.37.2-py3-none-any.whl (71 kB)
                              71.9/71.9 kB
4.6 MB/s eta 0:00:00
Downloading websockets-12.0-cp310-cp310-manylinux_2_5_x86_64.manylinux1_x8
6_64.manylinux_2_17_x86_64.manylinux2014_x86_64.whl (130 kB)
                              130.2/130.2 kB
3.1 MB/s eta 0:00:00
Installing collected packages: pydub, websockets, tomlkit, semantic-
version, ruff, python-multipart, orjson, h11, ffmpy, aiofiles, uvicorn,
starlette, httpcore, httpx, fastapi, gradio-client, gradio
  Attempting uninstall: tomlkit
    Found existing installation: tomlkit 0.13.0
    Uninstalling tomlkit-0.13.0:
      Successfully uninstalled tomlkit-0.13.0
Successfully installed aiofiles-23.2.1 fastapi-0.112.0 ffmpy-0.4.0 gradio-4.40.0
gradio-client-1.2.0 h11-0.14.0 httpcore-1.0.5 httpx-0.27.0 orjson-3.10.6
pydub-0.25.1 python-multipart-0.0.9 ruff-0.5.6 semantic-version-2.10.0
starlette-0.37.2 tomlkit-0.12.0 uvicorn-0.30.5 websockets-12.0
```

[4]: `!pip install datasets transformers torch faiss-cpu`

```
Collecting datasets
  Downloading datasets-2.20.0-py3-none-any.whl.metadata (19 kB)
Requirement already satisfied: transformers in /usr/local/lib/python3.10/dist-
packages (4.42.4)
Requirement already satisfied: torch in /usr/local/lib/python3.10/dist-packages
(2.3.1+cu121)
Collecting faiss-cpu
  Downloading faiss_cpu-1.8.0.post1-cp310-cp310-manylinux_2_17_x86_64.manylinux2
014_x86_64.whl.metadata (3.7 kB)
Requirement already satisfied: filelock in /usr/local/lib/python3.10/dist-
```

packages (from datasets) (3.15.4)
Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.10/dist-packages (from datasets) (1.26.4)
Collecting pyarrow>=15.0.0 (from datasets)
  Downloading pyarrow-17.0.0-cp310-cp310-manylinux_2_28_x86_64.whl.metadata (3.3 kB)
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.metadata (10 kB)
Requirement already satisfied: pandas in /usr/local/lib/python3.10/dist-packages (from datasets) (2.1.4)
Collecting requests>=2.32.2 (from datasets)
  Downloading requests-2.32.3-py3-none-any.whl.metadata (4.6 kB)
Requirement already satisfied: tqdm>=4.66.3 in /usr/local/lib/python3.10/dist-packages (from datasets) (4.66.4)
Collecting xxhash (from datasets)
  Downloading
xxhash-3.4.1-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (12 kB)
Collecting multiprocess (from datasets)
  Downloading multiprocess-0.70.16-py310-none-any.whl.metadata (7.2 kB)
Collecting fsspec<=2024.5.0,>=2023.1.0 (from
fsspec[http]<=2024.5.0,>=2023.1.0->datasets)
  Downloading fsspec-2024.5.0-py3-none-any.whl.metadata (11 kB)
Requirement already satisfied: aiohttp in /usr/local/lib/python3.10/dist-packages (from datasets) (3.10.0)
Requirement already satisfied: huggingface-hub>=0.21.2 in /usr/local/lib/python3.10/dist-packages (from datasets) (0.23.5)
Requirement already satisfied: packaging in /usr/local/lib/python3.10/dist-packages (from datasets) (24.1)
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.10/dist-packages (from datasets) (6.0.1)
Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/python3.10/dist-packages (from transformers) (2024.5.15)
Requirement already satisfied: safetensors>=0.4.1 in /usr/local/lib/python3.10/dist-packages (from transformers) (0.4.3)
Requirement already satisfied: tokenizers<0.20,>=0.19 in /usr/local/lib/python3.10/dist-packages (from transformers) (0.19.1)
Requirement already satisfied: typing-extensions>=4.8.0 in /usr/local/lib/python3.10/dist-packages (from torch) (4.12.2)
Requirement already satisfied: sympy in /usr/local/lib/python3.10/dist-packages (from torch) (1.13.1)
Requirement already satisfied: networkx in /usr/local/lib/python3.10/dist-packages (from torch) (3.3)
Requirement already satisfied: jinja2 in /usr/local/lib/python3.10/dist-packages (from torch) (3.1.4)
Requirement already satisfied: nvidia-cuda-nvrtc-cu12==12.1.105 in

/usr/local/lib/python3.10/dist-packages (from torch) (12.1.105)
Requirement already satisfied: nvidia-cuda-runtime-cu12==12.1.105 in
/usr/local/lib/python3.10/dist-packages (from torch) (12.1.105)
Requirement already satisfied: nvidia-cuda-cupti-cu12==12.1.105 in
/usr/local/lib/python3.10/dist-packages (from torch) (12.1.105)
Requirement already satisfied: nvidia-cudnn-cu12==8.9.2.26 in
/usr/local/lib/python3.10/dist-packages (from torch) (8.9.2.26)
Requirement already satisfied: nvidia-cublas-cu12==12.1.3.1 in
/usr/local/lib/python3.10/dist-packages (from torch) (12.1.3.1)
Requirement already satisfied: nvidia-cufft-cu12==11.0.2.54 in
/usr/local/lib/python3.10/dist-packages (from torch) (11.0.2.54)
Requirement already satisfied: nvidia-curand-cu12==10.3.2.106 in
/usr/local/lib/python3.10/dist-packages (from torch) (10.3.2.106)
Requirement already satisfied: nvidia-cusolver-cu12==11.4.5.107 in
/usr/local/lib/python3.10/dist-packages (from torch) (11.4.5.107)
Requirement already satisfied: nvidia-cusparse-cu12==12.1.0.106 in
/usr/local/lib/python3.10/dist-packages (from torch) (12.1.0.106)
Requirement already satisfied: nvidia-nccl-cu12==2.20.5 in
/usr/local/lib/python3.10/dist-packages (from torch) (2.20.5)
Requirement already satisfied: nvidia-nvtx-cu12==12.1.105 in
/usr/local/lib/python3.10/dist-packages (from torch) (12.1.105)
Requirement already satisfied: triton==2.3.1 in /usr/local/lib/python3.10/dist-
packages (from torch) (2.3.1)
Requirement already satisfied: nvidia-nvjitlink-cu12 in
/usr/local/lib/python3.10/dist-packages (from nvidia-cusolver-
cu12==11.4.5.107->torch) (12.6.20)
Requirement already satisfied: aiohappyeyeballs>=2.3.0 in
/usr/local/lib/python3.10/dist-packages (from aiohttp->datasets) (2.3.4)
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: charset-normalizer<4,>=2 in
/usr/local/lib/python3.10/dist-packages (from requests>=2.32.2->datasets)
(3.3.2)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-
packages (from requests>=2.32.2->datasets) (3.7)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.10/dist-packages (from requests>=2.32.2->datasets)
(2.0.7)

Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.10/dist-packages (from requests>=2.32.2->datasets)
(2024.7.4)
Requirement already satisfied: MarkupSafe>=2.0 in
/usr/local/lib/python3.10/dist-packages (from jinja2->torch) (2.1.5)
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) (2024.1)
Requirement already satisfied: tzdata>=2022.1 in /usr/local/lib/python3.10/dist-
packages (from pandas->datasets) (2024.1)
Requirement already satisfied: mpmath<1.4,>=1.1.0 in
/usr/local/lib/python3.10/dist-packages (from sympy->torch) (1.3.0)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-
packages (from python-dateutil>=2.8.2->pandas->datasets) (1.16.0)
Downloading datasets-2.20.0-py3-none-any.whl (547 kB)
                          547.8/547.8 kB
14.1 MB/s eta 0:00:00
Downloading
faiss_cpu-1.8.0.post1-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl
(27.0 MB)
                          27.0/27.0 MB
32.2 MB/s eta 0:00:00
Downloading dill-0.3.8-py3-none-any.whl (116 kB)
                          116.3/116.3 kB
5.0 MB/s eta 0:00:00
Downloading fsspec-2024.5.0-py3-none-any.whl (316 kB)
                          316.1/316.1 kB
17.1 MB/s eta 0:00:00
Downloading pyarrow-17.0.0-cp310-cp310-manylinux_2_28_x86_64.whl (39.9 MB)
                          39.9/39.9 MB
16.4 MB/s eta 0:00:00
Downloading requests-2.32.3-py3-none-any.whl (64 kB)
                          64.9/64.9 kB
3.4 MB/s eta 0:00:00
Downloading multiprocess-0.70.16-py310-none-any.whl (134 kB)
                          134.8/134.8 kB
8.1 MB/s eta 0:00:00
Downloading
xxhash-3.4.1-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (194 kB)
                          194.1/194.1 kB
6.6 MB/s eta 0:00:00
Installing collected packages: xxhash, requests, pyarrow, fsspec, faiss-
cpu, dill, multiprocess, datasets
  Attempting uninstall: requests
    Found existing installation: requests 2.31.0
    Uninstalling requests-2.31.0:
      Successfully uninstalled requests-2.31.0

```
      Attempting uninstall: pyarrow
        Found existing installation: pyarrow 14.0.2
        Uninstalling pyarrow-14.0.2:
          Successfully uninstalled pyarrow-14.0.2
      Attempting uninstall: fsspec
        Found existing installation: fsspec 2024.6.1
        Uninstalling fsspec-2024.6.1:
          Successfully uninstalled fsspec-2024.6.1
ERROR: pip's dependency resolver does not currently take into account all
the packages that are installed. This behaviour is the source of the following
dependency conflicts.
cudf-cu12 24.4.1 requires pyarrow<15.0.0a0,>=14.0.1, but you have pyarrow 17.0.0
which is incompatible.
gcsfs 2024.6.1 requires fsspec==2024.6.1, but you have fsspec 2024.5.0 which is
incompatible.
google-colab 1.0.0 requires requests==2.31.0, but you have requests 2.32.3 which
is incompatible.
ibis-framework 8.0.0 requires pyarrow<16,>=2, but you have pyarrow 17.0.0 which
is incompatible.
Successfully installed datasets-2.20.0 dill-0.3.8 faiss-cpu-1.8.0.post1
fsspec-2024.5.0 multiprocess-0.70.16 pyarrow-17.0.0 requests-2.32.3 xxhash-3.4.1
```

```python
[6]: !pip install --upgrade gradio
     import whisper
     from deep_translator import GoogleTranslator
     import torch
     from transformers import DPRQuestionEncoder, DPRContextEncoder,
       DPRQuestionEncoderTokenizer, DPRContextEncoderTokenizer
     import faiss
     import gradio as gr
```

```
Collecting gradio
  Downloading gradio-4.40.0-py3-none-any.whl.metadata (15 kB)
Collecting aiofiles<24.0,>=22.0 (from gradio)
  Downloading aiofiles-23.2.1-py3-none-any.whl.metadata (9.7 kB)
Requirement already satisfied: anyio<5.0,>=3.0 in
/usr/local/lib/python3.10/dist-packages (from gradio) (3.7.1)
Collecting fastapi (from gradio)
  Downloading fastapi-0.112.0-py3-none-any.whl.metadata (27 kB)
Collecting ffmpy (from gradio)
  Downloading ffmpy-0.4.0-py3-none-any.whl.metadata (2.9 kB)
Collecting gradio-client==1.2.0 (from gradio)
```

```
  Downloading gradio_client-1.2.0-py3-none-any.whl.metadata (7.1 kB)
Collecting httpx>=0.24.1 (from gradio)
  Downloading httpx-0.27.0-py3-none-any.whl.metadata (7.2 kB)
Requirement already satisfied: huggingface-hub>=0.19.3 in
/usr/local/lib/python3.10/dist-packages (from gradio) (0.23.5)
Requirement already satisfied: importlib-resources<7.0,>=1.3 in
/usr/local/lib/python3.10/dist-packages (from gradio) (6.4.0)
Requirement already satisfied: jinja2<4.0 in /usr/local/lib/python3.10/dist-
packages (from gradio) (3.1.4)
Requirement already satisfied: markupsafe~=2.0 in
/usr/local/lib/python3.10/dist-packages (from gradio) (2.1.5)
Requirement already satisfied: matplotlib~=3.0 in
/usr/local/lib/python3.10/dist-packages (from gradio) (3.7.1)
Requirement already satisfied: numpy<3.0,>=1.0 in
/usr/local/lib/python3.10/dist-packages (from gradio) (1.26.4)
Collecting orjson~=3.0 (from gradio)
  Downloading orjson-3.10.6-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_
64.whl.metadata (50 kB)
                         50.4/50.4 kB
2.9 MB/s eta 0:00:00
Requirement already satisfied: packaging in
/usr/local/lib/python3.10/dist-packages (from gradio) (24.1)
Requirement already satisfied: pandas<3.0,>=1.0 in
/usr/local/lib/python3.10/dist-packages (from gradio) (2.1.4)
Requirement already satisfied: pillow<11.0,>=8.0 in
/usr/local/lib/python3.10/dist-packages (from gradio) (9.4.0)
Requirement already satisfied: pydantic>=2.0 in /usr/local/lib/python3.10/dist-
packages (from gradio) (2.8.2)
Collecting pydub (from gradio)
  Downloading pydub-0.25.1-py2.py3-none-any.whl.metadata (1.4 kB)
Collecting python-multipart>=0.0.9 (from gradio)
  Downloading python_multipart-0.0.9-py3-none-any.whl.metadata (2.5 kB)
Requirement already satisfied: pyyaml<7.0,>=5.0 in
/usr/local/lib/python3.10/dist-packages (from gradio) (6.0.1)
Collecting ruff>=0.2.2 (from gradio)
  Downloading ruff-0.5.6-py3-none-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (24 kB)
Collecting semantic-version~=2.0 (from gradio)
  Downloading semantic_version-2.10.0-py2.py3-none-any.whl.metadata (9.7 kB)
Collecting tomlkit==0.12.0 (from gradio)
  Downloading tomlkit-0.12.0-py3-none-any.whl.metadata (2.7 kB)
Requirement already satisfied: typer<1.0,>=0.12 in
/usr/local/lib/python3.10/dist-packages (from gradio) (0.12.3)
Requirement already satisfied: typing-extensions~=4.0 in
/usr/local/lib/python3.10/dist-packages (from gradio) (4.12.2)
Requirement already satisfied: urllib3~=2.0 in /usr/local/lib/python3.10/dist-
packages (from gradio) (2.0.7)
Collecting uvicorn>=0.14.0 (from gradio)
```

```
  Downloading uvicorn-0.30.5-py3-none-any.whl.metadata (6.6 kB)
Requirement already satisfied: fsspec in /usr/local/lib/python3.10/dist-packages
(from gradio-client==1.2.0->gradio) (2024.5.0)
Collecting websockets<13.0,>=10.0 (from gradio-client==1.2.0->gradio)
  Downloading websockets-12.0-cp310-cp310-manylinux_2_5_x86_64.manylinux1_x86_64
.manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (6.6 kB)
Requirement already satisfied: idna>=2.8 in /usr/local/lib/python3.10/dist-
packages (from anyio<5.0,>=3.0->gradio) (3.7)
Requirement already satisfied: sniffio>=1.1 in /usr/local/lib/python3.10/dist-
packages (from anyio<5.0,>=3.0->gradio) (1.3.1)
Requirement already satisfied: exceptiongroup in /usr/local/lib/python3.10/dist-
packages (from anyio<5.0,>=3.0->gradio) (1.2.2)
Requirement already satisfied: certifi in /usr/local/lib/python3.10/dist-
packages (from httpx>=0.24.1->gradio) (2024.7.4)
Collecting httpcore==1.* (from httpx>=0.24.1->gradio)
  Downloading httpcore-1.0.5-py3-none-any.whl.metadata (20 kB)
Collecting h11<0.15,>=0.13 (from httpcore==1.*->httpx>=0.24.1->gradio)
  Downloading h11-0.14.0-py3-none-any.whl.metadata (8.2 kB)
Requirement already satisfied: filelock in /usr/local/lib/python3.10/dist-
packages (from huggingface-hub>=0.19.3->gradio) (3.15.4)
Requirement already satisfied: requests in /usr/local/lib/python3.10/dist-
packages (from huggingface-hub>=0.19.3->gradio) (2.32.3)
Requirement already satisfied: tqdm>=4.42.1 in /usr/local/lib/python3.10/dist-
packages (from huggingface-hub>=0.19.3->gradio) (4.66.4)
Requirement already satisfied: contourpy>=1.0.1 in
/usr/local/lib/python3.10/dist-packages (from matplotlib~=3.0->gradio) (1.2.1)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.10/dist-
packages (from matplotlib~=3.0->gradio) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in
/usr/local/lib/python3.10/dist-packages (from matplotlib~=3.0->gradio) (4.53.1)
Requirement already satisfied: kiwisolver>=1.0.1 in
/usr/local/lib/python3.10/dist-packages (from matplotlib~=3.0->gradio) (1.4.5)
Requirement already satisfied: pyparsing>=2.3.1 in
/usr/local/lib/python3.10/dist-packages (from matplotlib~=3.0->gradio) (3.1.2)
Requirement already satisfied: python-dateutil>=2.7 in
/usr/local/lib/python3.10/dist-packages (from matplotlib~=3.0->gradio) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-
packages (from pandas<3.0,>=1.0->gradio) (2024.1)
Requirement already satisfied: tzdata>=2022.1 in /usr/local/lib/python3.10/dist-
packages (from pandas<3.0,>=1.0->gradio) (2024.1)
Requirement already satisfied: annotated-types>=0.4.0 in
/usr/local/lib/python3.10/dist-packages (from pydantic>=2.0->gradio) (0.7.0)
Requirement already satisfied: pydantic-core==2.20.1 in
/usr/local/lib/python3.10/dist-packages (from pydantic>=2.0->gradio) (2.20.1)
Requirement already satisfied: click>=8.0.0 in /usr/local/lib/python3.10/dist-
packages (from typer<1.0,>=0.12->gradio) (8.1.7)
Requirement already satisfied: shellingham>=1.3.0 in
/usr/local/lib/python3.10/dist-packages (from typer<1.0,>=0.12->gradio) (1.5.4)
```

Requirement already satisfied: rich>=10.11.0 in /usr/local/lib/python3.10/dist-packages (from typer<1.0,>=0.12->gradio) (13.7.1)
Collecting starlette<0.38.0,>=0.37.2 (from fastapi->gradio)
  Downloading starlette-0.37.2-py3-none-any.whl.metadata (5.9 kB)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.7->matplotlib~=3.0->gradio) (1.16.0)
Requirement already satisfied: markdown-it-py>=2.2.0 in /usr/local/lib/python3.10/dist-packages (from rich>=10.11.0->typer<1.0,>=0.12->gradio) (3.0.0)
Requirement already satisfied: pygments<3.0.0,>=2.13.0 in /usr/local/lib/python3.10/dist-packages (from rich>=10.11.0->typer<1.0,>=0.12->gradio) (2.16.1)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests->huggingface-hub>=0.19.3->gradio) (3.3.2)
Requirement already satisfied: mdurl~=0.1 in /usr/local/lib/python3.10/dist-packages (from markdown-it-py>=2.2.0->rich>=10.11.0->typer<1.0,>=0.12->gradio) (0.1.2)
Downloading gradio-4.40.0-py3-none-any.whl (12.5 MB)
                        12.5/12.5 MB
83.5 MB/s eta 0:00:00
Downloading gradio_client-1.2.0-py3-none-any.whl (318 kB)
                        318.6/318.6 kB
19.6 MB/s eta 0:00:00
Downloading tomlkit-0.12.0-py3-none-any.whl (37 kB)
Downloading aiofiles-23.2.1-py3-none-any.whl (15 kB)
Downloading httpx-0.27.0-py3-none-any.whl (75 kB)
                        75.6/75.6 kB
4.5 MB/s eta 0:00:00
Downloading httpcore-1.0.5-py3-none-any.whl (77 kB)
                        77.9/77.9 kB
4.8 MB/s eta 0:00:00
Downloading
orjson-3.10.6-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (141 kB)
                        141.1/141.1 kB
9.3 MB/s eta 0:00:00
Downloading python_multipart-0.0.9-py3-none-any.whl (22 kB)
Downloading ruff-0.5.6-py3-none-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (10.2 MB)
                        10.2/10.2 MB
86.9 MB/s eta 0:00:00
Downloading semantic_version-2.10.0-py2.py3-none-any.whl (15 kB)
Downloading uvicorn-0.30.5-py3-none-any.whl (62 kB)
                        62.8/62.8 kB
4.1 MB/s eta 0:00:00
Downloading fastapi-0.112.0-py3-none-any.whl (93 kB)
                        93.1/93.1 kB

12

```
6.6 MB/s eta 0:00:00
Downloading ffmpy-0.4.0-py3-none-any.whl (5.8 kB)
Downloading pydub-0.25.1-py2.py3-none-any.whl (32 kB)
Downloading h11-0.14.0-py3-none-any.whl (58 kB)
                        58.3/58.3 kB
3.0 MB/s eta 0:00:00
Downloading starlette-0.37.2-py3-none-any.whl (71 kB)
                        71.9/71.9 kB
4.6 MB/s eta 0:00:00
Downloading websockets-12.0-cp310-cp310-manylinux_2_5_x86_64.manylinux1_x8
6_64.manylinux_2_17_x86_64.manylinux2014_x86_64.whl (130 kB)
                        130.2/130.2 kB
8.9 MB/s eta 0:00:00
Installing collected packages: pydub, websockets, tomlkit, semantic-
version, ruff, python-multipart, orjson, h11, ffmpy, aiofiles, uvicorn,
starlette, httpcore, httpx, fastapi, gradio-client, gradio
  Attempting uninstall: tomlkit
    Found existing installation: tomlkit 0.13.0
    Uninstalling tomlkit-0.13.0:
      Successfully uninstalled tomlkit-0.13.0
Successfully installed aiofiles-23.2.1 fastapi-0.112.0 ffmpy-0.4.0 gradio-4.40.0
gradio-client-1.2.0 h11-0.14.0 httpcore-1.0.5 httpx-0.27.0 orjson-3.10.6
pydub-0.25.1 python-multipart-0.0.9 ruff-0.5.6 semantic-version-2.10.0
starlette-0.37.2 tomlkit-0.12.0 uvicorn-0.30.5 websockets-12.0
```

[7]:
```python
whisper_model = whisper.load_model("base")
```

```
100%|                              | 139M/139M [00:02<00:00, 58.1MiB/s]
```

[8]:
```python
question_encoder = DPRQuestionEncoder.from_pretrained("facebook/
  ↪dpr-question_encoder-multiset-base")
context_encoder = DPRContextEncoder.from_pretrained("facebook/
  ↪dpr-ctx_encoder-multiset-base")
question_encoder_tokenizer = DPRQuestionEncoderTokenizer.
  ↪from_pretrained("facebook/dpr-question_encoder-multiset-base")
context_encoder_tokenizer = DPRContextEncoderTokenizer.
  ↪from_pretrained("facebook/dpr-ctx_encoder-multiset-base")

MAX_LENGTH = 256
```

```
/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://huggingface.co/settings/tokens), set it as secret in your Google Colab
and restart your session.
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:    0%|              | 0.00/493 [00:00<?, ?B/s]

pytorch_model.bin:    0%|              | 0.00/438M [00:00<?, ?B/s]
```

Some weights of the model checkpoint at facebook/dpr-question_encoder-multiset-
base were not used when initializing DPRQuestionEncoder:
['question_encoder.bert_model.pooler.dense.bias',
'question_encoder.bert_model.pooler.dense.weight']
- This IS expected if you are initializing DPRQuestionEncoder from the
checkpoint of a model trained on another task or with another architecture (e.g.
initializing a BertForSequenceClassification model from a BertForPreTraining
model).
- This IS NOT expected if you are initializing DPRQuestionEncoder from the
checkpoint of a model that you expect to be exactly identical (initializing a
BertForSequenceClassification model from a BertForSequenceClassification model).

```
config.json:    0%|              | 0.00/492 [00:00<?, ?B/s]

pytorch_model.bin:    0%|              | 0.00/438M [00:00<?, ?B/s]
```

Some weights of the model checkpoint at facebook/dpr-ctx_encoder-multiset-base
were not used when initializing DPRContextEncoder:
['ctx_encoder.bert_model.pooler.dense.bias',
'ctx_encoder.bert_model.pooler.dense.weight']
- This IS expected if you are initializing DPRContextEncoder from the checkpoint
of a model trained on another task or with another architecture (e.g.
initializing a BertForSequenceClassification model from a BertForPreTraining
model).
- This IS NOT expected if you are initializing DPRContextEncoder from the
checkpoint of a model that you expect to be exactly identical (initializing a
BertForSequenceClassification model from a BertForSequenceClassification model).

```
tokenizer_config.json:    0%|              | 0.00/28.0 [00:00<?, ?B/s]

vocab.txt:    0%|              | 0.00/232k [00:00<?, ?B/s]

tokenizer.json:    0%|              | 0.00/466k [00:00<?, ?B/s]

tokenizer_config.json:    0%|              | 0.00/28.0 [00:00<?, ?B/s]

vocab.txt:    0%|              | 0.00/232k [00:00<?, ?B/s]

tokenizer.json:    0%|              | 0.00/466k [00:00<?, ?B/s]
```

The tokenizer class you load from this checkpoint is not the same type as the
class this function is called from. It may result in unexpected tokenization.
The tokenizer class you load from this checkpoint is
'DPRQuestionEncoderTokenizer'.
The class this function is called from is 'DPRContextEncoderTokenizer'.

```
[9]: documents = [
         "This is Ananya",
          "Multilingual speech recognition allows systems to understand and process␣
     ↪spoken language in multiple languages."
          "This technology utilizes machine learning models that are trained on␣
     ↪extensive datasets containing speech samples from various languages to␣
     ↪achieve precise transcription and translation of spoken content."
          "A major difficulty in multilingual speech recognition is managing the␣
     ↪diverse range of accents and dialects within each language. Additionally,␣
     ↪factors like noisy surroundings and overlapping speech can greatly affect␣
     ↪the precision of speech recognition systems."
     ]
```

```
[10]: def encode_contexts(documents):
          inputs = context_encoder_tokenizer(documents, return_tensors="pt",␣
     ↪padding=True, truncation=True, max_length=MAX_LENGTH)
          with torch.no_grad():
              embeddings = context_encoder(**inputs).pooler_output
          return embeddings
```

```
[11]: def create_faiss_index(embeddings):
          dimension = embeddings.shape[1]
          index = faiss.IndexFlatL2(dimension)
          index.add(embeddings.numpy())
          return index

      context_embeddings = encode_contexts(documents)
      faiss_index = create_faiss_index(context_embeddings)
```

```
[12]: def retrieve_answer(question, faiss_index, documents):
          question_input = question_encoder_tokenizer(question, return_tensors="pt",␣
     ↪truncation=True, padding=True, max_length=MAX_LENGTH)
          with torch.no_grad():
              question_embedding = question_encoder(**question_input).pooler_output
          distances, indices = faiss_index.search(question_embedding.numpy(), k=1)  #␣
     ↪k=1 for top-1 result
          most_similar_doc_index = indices[0][0]
          return documents[most_similar_doc_index]
```

```
[13]: def translate_with_deep_translator(text, source_language='auto',␣
     ↪target_language='en'):
          translator = GoogleTranslator(source=source_language,␣
     ↪target=target_language)
          return translator.translate(text)
```

```python
[14]: def transcribe_translate_query(audio):
          audio = whisper.load_audio(audio)
          audio = whisper.pad_or_trim(audio)
          mel = whisper.log_mel_spectrogram(audio).to(whisper_model.device)
          _, probs = whisper_model.detect_language(mel)
          detected_language = max(probs, key=probs.get)
          options = whisper.DecodingOptions()
          result = whisper.decode(whisper_model, mel, options)
          transcription = result.text
          translated_text = translate_with_deep_translator(transcription,
      ↪source_language=detected_language, target_language='en')
          return detected_language, transcription, translated_text
```

```python
[15]: def process_audio(audio):
          detected_language, transcription, translated_text =
      ↪transcribe_translate_query(audio)
          rag_answer = retrieve_answer(translated_text, faiss_index, documents)
          return rag_answer if rag_answer else translated_text
```

```python
[16]: gr.Interface(
          fn=process_audio,
          inputs=gr.Audio(type="filepath"),
          outputs=gr.Textbox(),
          live=True,
          title='Multilingual Speech Recognition With RAG'
      ).launch()
```

Setting queue=True in a Colab notebook requires sharing enabled. Setting
`share=True` (you can turn this off by setting `share=False` in `launch()`
explicitly).

Colab notebook detected. To show errors in colab notebook, set debug=True in
launch()
Running on public URL: https://1de70327701d256440.gradio.live

This share link expires in 72 hours. For free permanent hosting and GPU
upgrades, run `gradio deploy` from Terminal to deploy to Spaces
(https://huggingface.co/spaces)

<IPython.core.display.HTML object>

[16]:

```python
[17]: !pip install jiwer
```

Collecting jiwer
  Downloading jiwer-3.0.4-py3-none-any.whl.metadata (2.6 kB)
Requirement already satisfied: click<9.0.0,>=8.1.3 in

```
/usr/local/lib/python3.10/dist-packages (from jiwer) (8.1.7)
Collecting rapidfuzz<4,>=3 (from jiwer)
  Downloading rapidfuzz-3.9.5-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x8
6_64.whl.metadata (12 kB)
Downloading jiwer-3.0.4-py3-none-any.whl (21 kB)
Downloading
rapidfuzz-3.9.5-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (3.4
MB)
                            3.4/3.4 MB
41.8 MB/s eta 0:00:00
Installing collected packages: rapidfuzz, jiwer
Successfully installed jiwer-3.0.4 rapidfuzz-3.9.5
```

[18]:
```python
from jiwer import wer

# Replace with your actual transcriptions
reference_transcriptions = ["This is a test sentence."]
hypothesis_transcriptions = ["This is test sentence."]

# Compute WER
wer_score = wer(reference_transcriptions, hypothesis_transcriptions)
print(f"Word Error Rate (WER): {wer_score:.4f}")
```

```
Word Error Rate (WER): 0.2000
```

[19]:
```python
!pip install sacrebleu
```

```
Collecting sacrebleu
  Downloading sacrebleu-2.4.2-py3-none-any.whl.metadata (58 kB)
                              58.0/58.0 kB
3.4 MB/s eta 0:00:00
Collecting portalocker (from sacrebleu)
  Downloading portalocker-2.10.1-py3-none-any.whl.metadata (8.5 kB)
Requirement already satisfied: regex in /usr/local/lib/python3.10/dist-packages
(from sacrebleu) (2024.5.15)
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) (1.26.4)
Collecting colorama (from sacrebleu)
  Downloading colorama-0.4.6-py2.py3-none-any.whl.metadata (17 kB)
Requirement already satisfied: lxml in /usr/local/lib/python3.10/dist-packages
(from sacrebleu) (4.9.4)
Downloading sacrebleu-2.4.2-py3-none-any.whl (106 kB)
                             106.7/106.7 kB
7.6 MB/s eta 0:00:00
Downloading colorama-0.4.6-py2.py3-none-any.whl (25 kB)
Downloading portalocker-2.10.1-py3-none-any.whl (18 kB)
```

```
Installing collected packages: portalocker, colorama, sacrebleu
Successfully installed colorama-0.4.6 portalocker-2.10.1 sacrebleu-2.4.2
```

```python
from nltk.translate.bleu_score import corpus_bleu

# Example sentences
references = [[["this", "is", "a", "test"]]]  # Note: nested list for references
hypothesis = [["this", "is", "a", "test"]]  # Note: nested list for hypothesis

# Calculate BLEU score
score = corpus_bleu(references, hypothesis)
print(f"BLEU Score: {score:.2f}")
```

```
BLEU Score: 1.00
```