

```
!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
```

```
!pip install deep-translator
```

```
→ Requirement already satisfied: deep-translator in /usr/local/lib/python3.10/dist-packages (1.11.4)  
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) (2.32.3)  
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)  
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.  
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
```

```
!pip install --upgrade gradio
```

```
→ Requirement already satisfied: gradio in /usr/local/lib/python3.10/dist-packages (4.40.0)  
Requirement already satisfied: aiofiles<24.0,>=22.0 in /usr/local/lib/python3.10/dist-packages (from gradio) (23.2.1)  
Requirement already satisfied: anyio<5.0,>=3.0 in /usr/local/lib/python3.10/dist-packages (from gradio) (3.7.1)  
Requirement already satisfied: fastapi in /usr/local/lib/python3.10/dist-packages (from gradio) (0.112.0)  
Requirement already satisfied: ffmpy in /usr/local/lib/python3.10/dist-packages (from gradio) (0.4.0)  
Requirement already satisfied: gradio-client==1.2.0 in /usr/local/lib/python3.10/dist-packages (from gradio) (1.2.0)  
Requirement already satisfied: httpx=>0.24.1 in /usr/local/lib/python3.10/dist-packages (from gradio) (0.27.0)  
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)  
Requirement already satisfied: orjson~3.0 in /usr/local/lib/python3.10/dist-packages (from gradio) (3.10.6)  
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)  
Requirement already satisfied: pydub in /usr/local/lib/python3.10/dist-packages (from gradio) (0.25.1)  
Requirement already satisfied: python-multipart=>0.0.9 in /usr/local/lib/python3.10/dist-packages (from gradio) (0.0.9)  
Requirement already satisfied: pyyaml<7.0,>=5.0 in /usr/local/lib/python3.10/dist-packages (from gradio) (6.0.1)  
Requirement already satisfied: ruff=>0.2.2 in /usr/local/lib/python3.10/dist-packages (from gradio) (0.5.6)  
Requirement already satisfied: semantic-version~2.0 in /usr/local/lib/python3.10/dist-packages (from gradio) (2.10.0)  
Requirement already satisfied: tomlkit==0.12.0 in /usr/local/lib/python3.10/dist-packages (from gradio) (0.12.0)  
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)  
Requirement already satisfied: uvicorn=>0.14.0 in /usr/local/lib/python3.10/dist-packages (from gradio) (0.30.5)  
Requirement already satisfied: fsspec in /usr/local/lib/python3.10/dist-packages (from gradio-client==1.2.0->gradio) (2024.5.0)  
Requirement already satisfied: websockets<13.0,>=10.0 in /usr/local/lib/python3.10/dist-packages (from gradio-client==1.2.0->gradio) (12.0)  
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)  
Requirement already satisfied: httpcore==1.* in /usr/local/lib/python3.10/dist-packages (from httpx=>0.24.1->gradio) (1.0.5)  
Requirement already satisfied: h11<0.15,>=0.13 in /usr/local/lib/python3.10/dist-packages (from httpcore==1.*->httpx=>0.24.1->gradio) (0.14.0)  
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: cyclers=>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)  
Requirement already satisfied: starlette<0.38.0,>=0.37.2 in /usr/local/lib/python3.10/dist-packages (from fastapi->gradio) (0.37.2)  
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)  
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)  
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests->huggingface-hub=>0.19.3->gra
```

```
!pip install datasets transformers torch faiss-cpu
```

```
→ Requirement already satisfied: datasets in /usr/local/lib/python3.10/dist-packages (2.20.0)  
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)  
Requirement already satisfied: faiss-cpu in /usr/local/lib/python3.10/dist-packages (1.8.0.post1)  
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)  
Requirement already satisfied: pyarrow=>15.0.0 in /usr/local/lib/python3.10/dist-packages (from datasets) (17.0.0)  
Requirement already satisfied: pyarrow-hotfix in /usr/local/lib/python3.10/dist-packages (from datasets) (0.6)  
Requirement already satisfied: dill<0.3.9,>=0.3.0 in /usr/local/lib/python3.10/dist-packages (from datasets) (0.3.8)  
Requirement already satisfied: pandas in /usr/local/lib/python3.10/dist-packages (from datasets) (2.1.4)  
Requirement already satisfied: requests=>2.32.2 in /usr/local/lib/python3.10/dist-packages (from datasets) (2.32.3)
```

```
Requirement already satisfied: tqdm>=4.66.3 in /usr/local/lib/python3.10/dist-packages (from datasets) (4.66.4)
Requirement already satisfied: xxhash in /usr/local/lib/python3.10/dist-packages (from datasets) (3.4.1)
Requirement already satisfied: multiprocess in /usr/local/lib/python3.10/dist-packages (from datasets) (0.70.16)
Requirement already satisfied: fsspec<=2024.5.0,>=2023.1.0 in /usr/local/lib/python3.10/dist-packages (from fsspec[http]<=2024.5.0,>=2023.1.0->dat
Requirement already satisfied: aiohttp in /usr/local/lib/python3.10/dist-packages (from datasets) (3.9.5)
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
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)
```

```
import whisper
from deep_translator import GoogleTranslator
import torch
from transformers import DPRQuestionEncoder, DPRContextEncoder, DPRQuestionEncoderTokenizer, DPRContextEncoderTokenizer
import faiss
import gradio as gr

whisper_model = whisper.load_model("base")

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

→ Some weights of the model checkpoint at facebook/dpr-question_encoder-multiset-base were not used when initializing DPRQuestionEncoder: ['question
- This IS expected if you are initializing DPRQuestionEncoder from the checkpoint of a model trained on another task or with another architecture
- This IS NOT expected if you are initializing DPRQuestionEncoder from the checkpoint of a model that you expect to be exactly identical (initiali
Some weights of the model checkpoint at facebook/dpr-ctx_encoder-multiset-base were not used when initializing DPRContextEncoder: ['ctx_encoder.be
- This IS expected if you are initializing DPRContextEncoder from the checkpoint of a model trained on another task or with another architecture (
- This IS NOT expected if you are initializing DPRContextEncoder from the checkpoint of a model that you expect to be exactly identical (initiali
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 toke
The tokenizer class you load from this checkpoint is 'DPRQuestionEncoderTokenizer'.
The class this function is called from is 'DPRContextEncoderTokenizer'.
```

documents = [
 "This project is done by Yochana",
 "Multilingual speech recognition technology enables systems to understand and process spoken language in multiple languages. This technology uses
 "Multilingual speech recognition is widely used in virtual assistants like Siri and Alexa, customer service chatbots, and real-time translation ap
 "One of the main challenges in multilingual speech recognition is handling the wide variety of accents and dialects within each language. Addition
 "Future trends in multilingual speech recognition include advancements in contextual understanding and the ability to switch languages seamlessly
]

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

```

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)

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]

def translate_with_deep_translator(text, source_language='auto', target_language='en'):
    translator = GoogleTranslator(source=source_language, target=target_language)
    return translator.translate(text)

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

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

```

```

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()`)

Colab notebook detected. To show errors in colab notebook, set debug=True in launch()
Running on public URL: <https://dd0477d290dddeb285.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://h>

Multilingual Speech Recognition With RAG

The interface consists of two main sections. On the left is an audio player with a waveform visualization, current time (0:00), end time (0:04), and playback controls (volume, speed, and navigation arrows). Below the player is a 'Clear' button. On the right is a text area labeled 'output' containing the following text:

Future trends in multilingual speech recognition include advancements in contextual understanding and the ability to switch languages seamlessly during conversations. Researchers are also focusing on improving speech recognition for low-resource languages that lack large training datasets.

Below the output text is a 'Flag' button.

```
!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_x86_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 13.3 MB/s eta 0:00:00
Installing collected packages: rapidfuzz, jiwer
Successfully installed jiwer-3.0.4 rapidfuzz-3.9.5
```

```
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
```

```
!pip install sacrebleu
```

```
→ Collecting sacrebleu
  Downloading sacrebleu-2.4.2-py3-none-any.whl.metadata (58 kB)
  ━━━━━━━━━━━━━━━━ 58.0/58.0 kB 941.3 kB/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 1.8 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
```

```
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
```