

Building a Multilingual Speech Recognition Model for RAG Without Training

INTRODUCTION:

In this project , I have developed a multilingual speech recognition model for RAG without training , specifically Multilingual Whisper ,to enhance the capabilities of RAG to handling tasks across multiple language and a chatbot provide relevant responses. The application aims to preform speech recognition , translation and summarization of audio and video inputs.

OBJECTIVE:

To develop a multilingual speech recognition system use the pre-trained whisper model . Integrating the system with RAG to enable the chatbot to understand and respond to audio queries. It handle the audio or video in multiple language and converted into English to perform RAG .It uses gemma 2b LLM model for summarization .

METHODOLOGY:

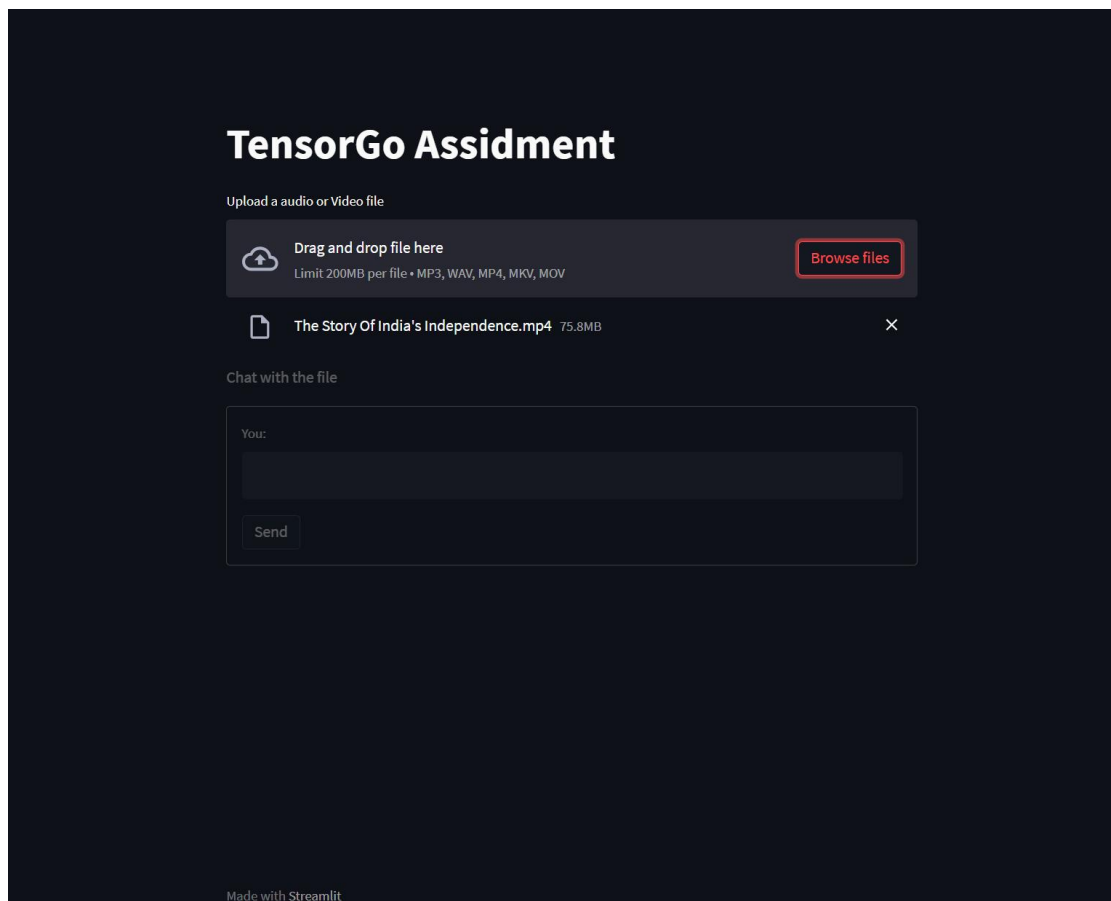
The Whisper v3 is a pre-trained model that can convert multilingual speech to text . Then the text is embedded in to vectors to perform RAG . So the gemma-2b-it model generate the summary of the speech . The RAG retrieve the relevant information from the vectors then its pass to the LLM model and generate the relevant answers for the question asked .

SYSTEM REQUIREMENTS :

- RAM: 8 GB or higher
- Processor: Intel Core i5 12th Gen or higher
- GPU: NVIDIA RTX 3050 with 8 GB VRAM or higher
- Operating System: Windows

IMPLEMENTATION:

- The source code for this application can be clone through git hub the link : <https://github.com/Mridul1321/Multilingual-speech-recognition> .
- To run this application we cuda setup should be done and the GPU should be NVIDIA and the VRAM should be greater then 8GB
- After downloading install the requirement.txt file in the environment where the cuda setup is done “**pip install -r requirements.txt**” .
- Once its done then you can run the application in the terminal by using the command : “**streamlit run app.py**”.
- Now the application will running on your localhost ,Which is demonstrated in the following figure



- Now the audio or the video file can be loaded into the application
- Once the file is loaded it recognize the text in the speech and translate the form other languages to English .

- Then the text is embedded into vectors and the text is passed to the gemma-2b-it model and it provides the summary of the speech

TensorGo Assidment

Chat with the file

You:

Send

Bot: The Summery of the file : Sure, here's a summary of the text:

The British came to India for trading and became the major force in India with the establishment of their offices and forts. The British Raj in India deprived the Indians of basic needs, socially and economically and legally discriminated against, misbehaved, mistreated and tortured. However, the Indian soldiers fought bravely in the First World War and World War I, and after extensive struggles, India finally gained its independence on August 15, 1947.

Bot: The content extracted form the audio file :

The British came to India for trading. They made their offices and forts in various parts of India But eventually the British East India Company became the major force in India the company's troops led by Robert Clive defeated Siraj Udulla the ruler of Bengal in 1757 in the Battle of Palasi and that was the beginning of British rule known as British Raj in India. Under the British rule, the Indians were deprived of basic needs, socially and economically and legally discriminated against, misbehaved, mistreated and tortured. India's first war of independence was a revolt of Indian soldiers and rulers against British rule.

- Once its done then you can pass your question then its performs RAG operation, retrieves the relevant answers form the speech and pass to LLM gemma-2b-it model then is generate the response .

TensorGo Assidment

Chat with the file

You:

Send

Bot: The passage mentions the British came to India for trading. However, it does not explicitly state the purpose of their visit or why they came to India.

You: For what perpose the british came to india the when the came

Bot: The Summery of the file : Sure, here's a summary of the text:

The British came to India for trading and became the major force in India with the establishment of their offices and forts. The British Raj in India deprived the Indians of basic needs, socially and economically and legally discriminated against, misbehaved, mistreated and tortured. However, the Indian soldiers fought bravely in the First World War and World War I, and after extensive struggles, India finally gained its independence on August 15, 1947.

Bot: The content extracted form the audio file :

The British came to India for trading. They made their offices and forts in various parts of India But eventually the British East India Company became the major force in India the company's troops led by Robert Clive defeated Siraj Udulla the ruler of Bengal in 1757 in the Battle of Palasi and that was the

- Finally the audio or the video gets converted to text and translated to English then the LLM model provides the summary of the speech and also answers the questions using RAG.
- The sample demo video can be accessed through the link:
https://drive.google.com/drive/folders/1-PPN__0hmG-BbU55w9wg2htvXt-T9IIK

CONCLUSION:

Finally the project developed successfully by the Multilingual Whisper model for speech recognition and RAG for translation and summarization of audio and video inputs. The chatbot can enhance the application to handle the queries regarding the speech .