Techshila’24 ML Report

Submission by: Himalaya Bhawan

Team Members:
Aishwarya Ahuja
Palak Jhamnani
Paridhi Jain,
Shruti Chouhan
Shoilayee Chaudhuri,
Vidhi Patidar

PROBLEM OVERVIEW:

We tried to build an AI-based interviewer, which asks topic-specific questions and analyses the user’s audio answer and gives feedback on it. It also corrects any grammatical errors and provides a reference-answer to the user.

DATASET CREATION:

Text data-set creation:

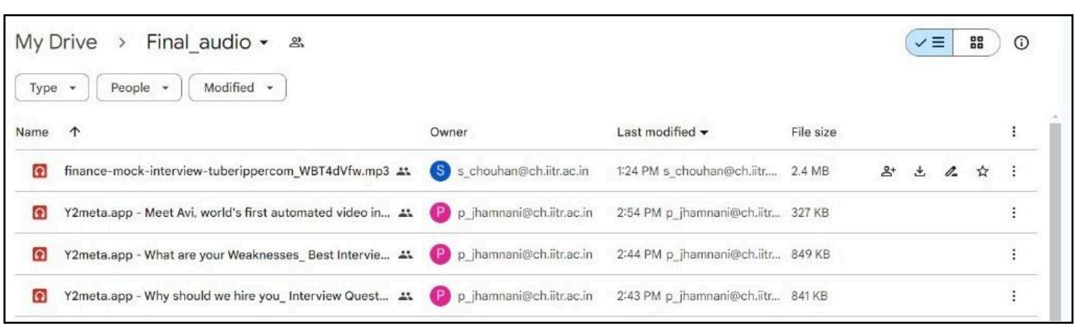
We scraped pdfs and websites containing interview question answers of various domains and created a dataset with 3048 data points, of the following format:

```
In [20]: import pandas as pd
         df = pd.read_csv("C:\\Users\\Admin\\Desktop\\techshila dataset\\final6.csv")
         df.head()
Out[20]:
```

	Question	Answer	Prompt
0	Tell me about yourself.	If there's one question you're bound to encoun...	Give me consultancy questions for interview
1	Why do you want to work for our company?	Before the interview, research the types of pr...	Give me consultancy questions for interview
2	Why consulting?	Consulting is a highly competitive field, and ...	Give me consultancy questions for interview
3	Why are you the best person for this position?...	This question is similar to an elevator pitch....	Give me consultancy questions for interview
4	What is your leadership style?	This question allows the interviewer to assess...	Give me consultancy questions for interview

Audio data-set:

We extracted the audio from the url of 40 youtube videos on live interviews. They cover various topics and speaking styles in the English language. They were then converted to mp3 format and stored in a folder on gdrive.



MODEL OVERVIEW:

Text Model Development:

We explored various text-generation models (CausalLM) like Alpaca Large, Mistralai(Mixtral-8x7B), Llama and Google Gemma. But due to our computational limitations and the big size of the models, we opted for Llama-2-7b-chat-hf.

This choice was driven by its

1. Demonstrably good performance
2. Efficient training times
3. Ability to deliver results quickly.

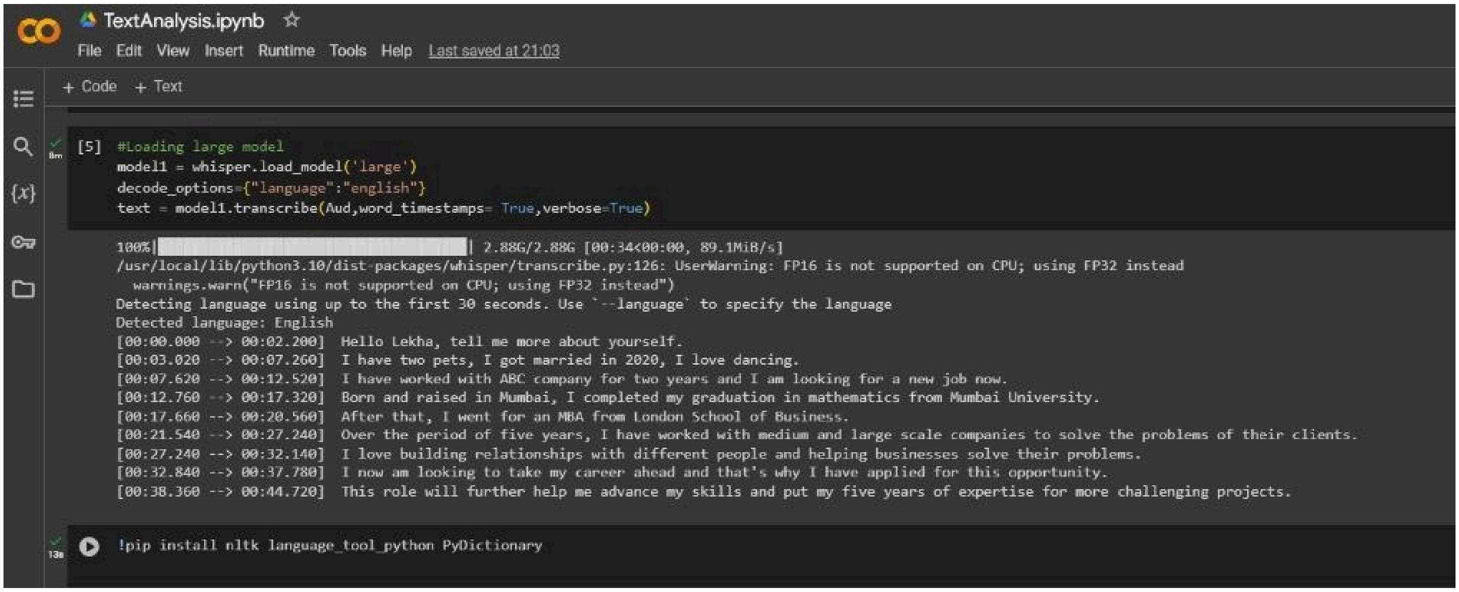
Audio Model Development:

Audio processing:

The script imports necessary libraries including ‘os’, ‘tqdm’, ‘torch’, ‘librosa’, and ‘gradio’ which are used for various tasks such as file operation, progress tracking, numerical computations, audio processing, and visualization. First, launched gradio with a wav2vec2 model of hugging face to record the audio. Then loaded a pre-trained Whisper model, it is a speech recognition model, and 'base' refers to a large pre-trained model variant. This model is used to transcribe audio files later in the script. Calculated the speaking pace (words per minute) based on the duration of the audio file. After processing all audio files, it prints the collected speaking paces and gives feedback to the user after comparing it with the ideal speaking pace, i.e., 140-160 (wpm).

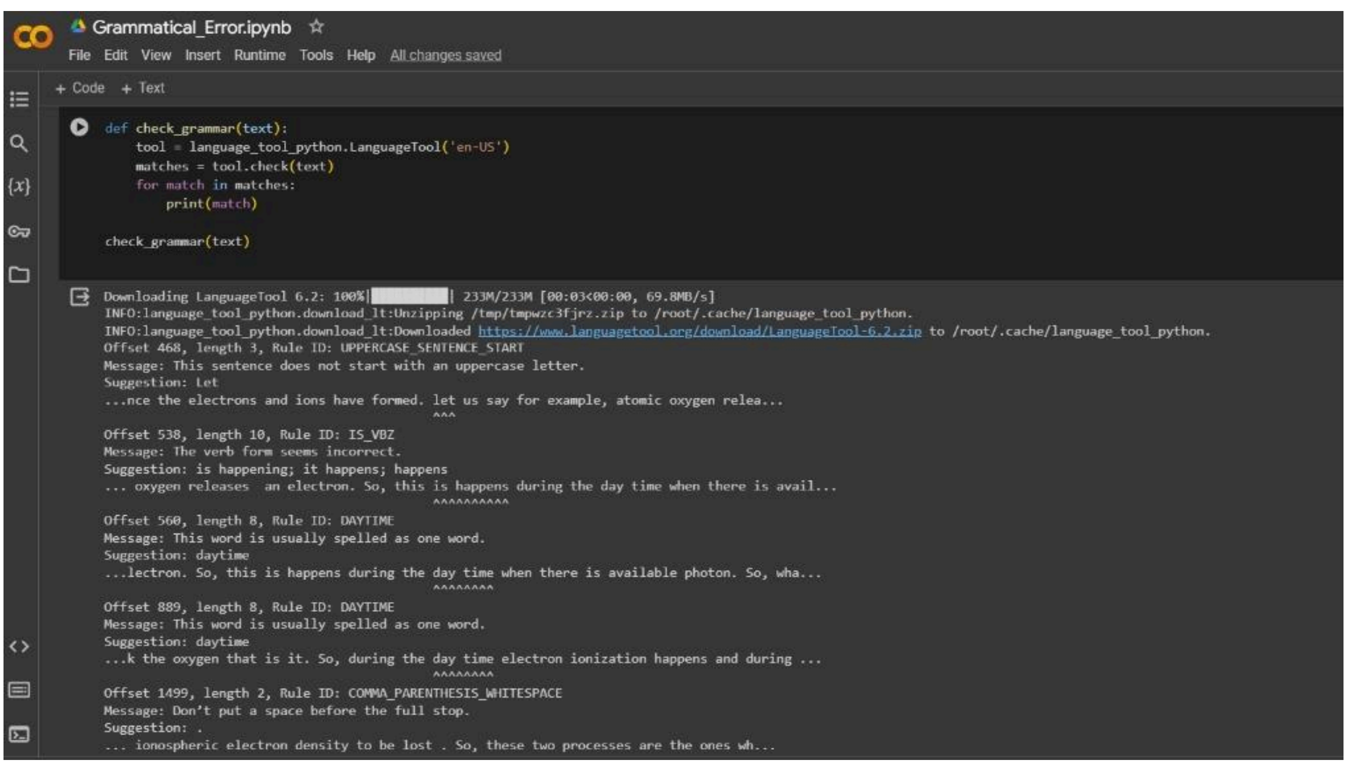
Speech-to-Text Conversion

We utilized Whisper, a state-of-the-art speech-to-text conversion tool, to transcribe spoken responses into text format. Whisper offers high accuracy and robustness, making it suitable for our project's requirements. This step was crucial in enabling the system to process spoken inputs for further analysis



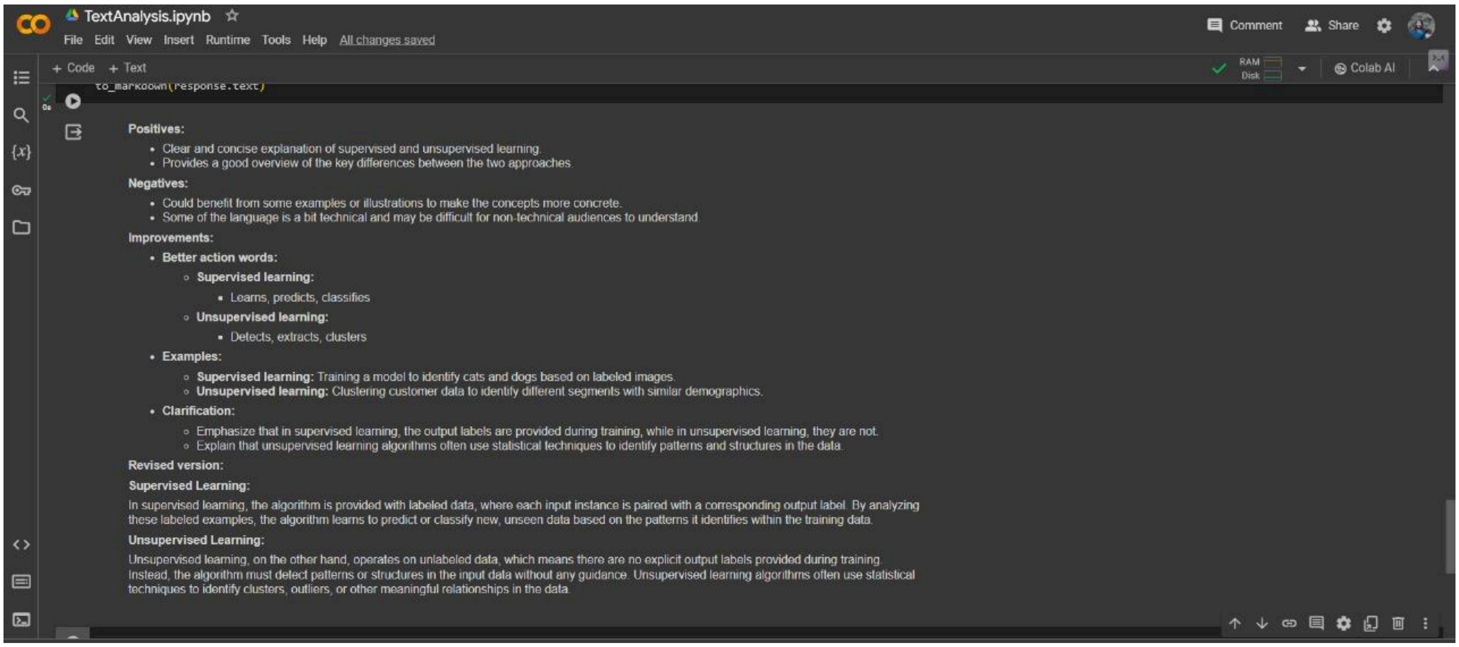
Grammatical Error Detection with LanguageTool

After converting speech to text, we employed LanguageTool, a powerful grammar checking tool, to identify grammatical errors within the transcribed text. LanguageTool utilizes advanced algorithms to detect various types of grammatical mistakes, including punctuation errors, spelling errors, and syntactical inconsistencies. By integrating LanguageTool into our pipeline, we ensured the user gets accurate and grammatically correct responses.



Integration with Gemini AI for Feedback Generation

To provide meaningful feedback to users, we integrated Gemini AI, an advanced natural language processing engine, into our system. Gemini AI analyzes the transcribed text, identifies strengths and weaknesses in the user's response, and generates personalized feedback. By leveraging Gemini AI's capabilities, we were able to offer actionable insights for improvement.



The feedback includes the following components:

1. Positives: Highlighting strengths or positive aspects of the user's response.
2. Negatives: Identifying areas for improvement or weaknesses in the response.
3. Suggestions for Improvement: Offering constructive suggestions or recommendations to enhance the quality of the response.
4. Action Words: Recommending impactful or persuasive words that users can incorporate into their responses to enhance clarity and effectiveness in interviews

METRICS:

BLEU Score: 0.47
F1 Score: 0.20

CHALLENGES FACED:

1. The creation of the datasets was a time-consuming process, requiring multiple iterations to ensure its accuracy and compatibility with the model.

2. We tried to explore many CausalLM models but was hampered by constrained computational resources, resulting in repeated memory errors. However, Llama-2-7b-chat-hf worked out for us and we used it for the model.
3. The implementation of audio libraries presented complexities, especially in achieving consistent recording quality.

LIMITATIONS AND IMPROVEMENTS:

1. We were supposed to ensure conversational depth, but we could not integrate it with our existing model. More time would perhaps let us work in this direction and make improvements.

2. While analyzing the input audio of the user, we aimed at real-time feedback but that required more sophistication which in turn required much more research and exploration, which could not be managed.