(TR-103) PROMPT ENGINEERING –

Training Day 7 Report:

Speech-to-Text Conversion and Integration with Gemini

Objective:

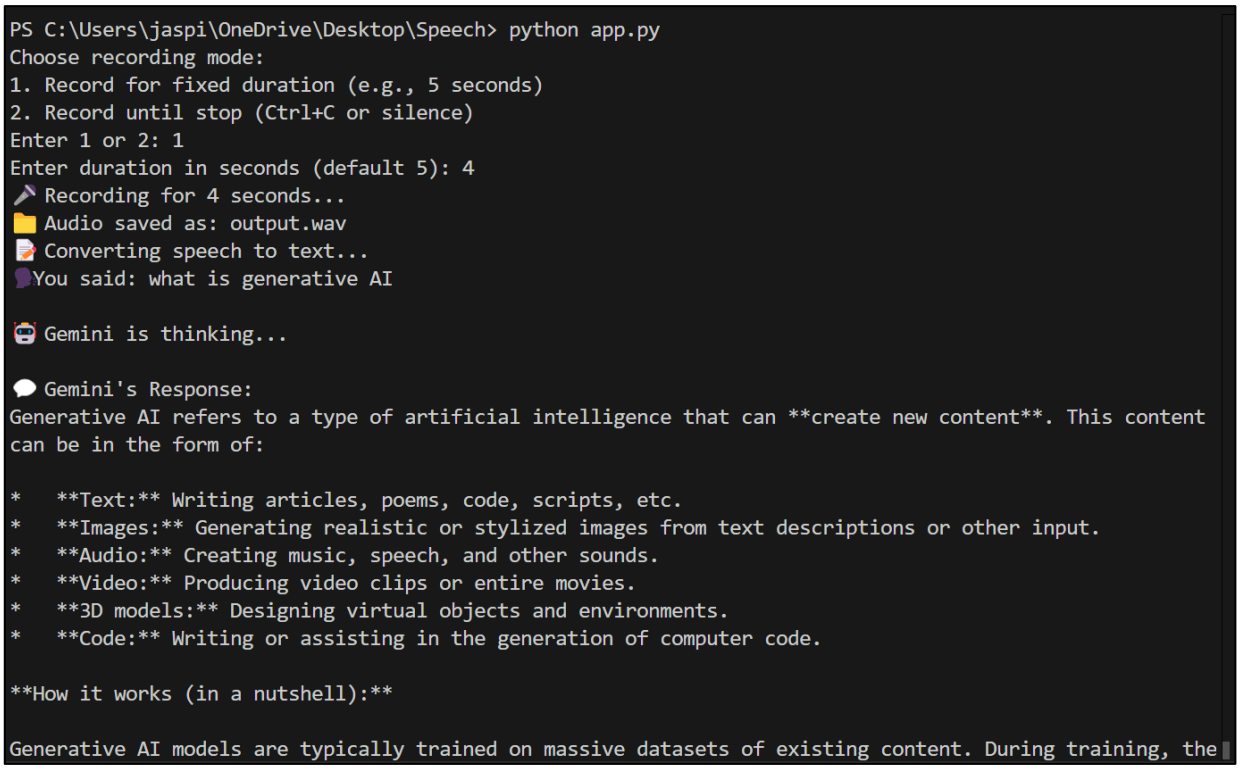
The goal for the day was to implement and test a speech-to-text system using Python and integrate it with the Gemini Generative AI model to get intelligent responses based on user input through voice.

Work Done:

- Utilized the speech_recognition Python library to capture audio from the microphone.
- Implemented two modes:
  - Fixed Duration Recording (e.g., 5 seconds)
  - Recording Until Silence or Manual Stop
- Added functionality to save the audio as a .wav file for backup or analysis.
- Used Google's Speech Recognition API to convert the captured audio into text.
- Configured and used the Gemini Flash Lite model (models/gemini-2.0-flash-lite) to process and generate responses based on the transcribed text.
- Integrated a function to send the converted text to the Gemini model and display the AI-generated response to the user.

Output:

- Successfully captured and transcribed voice inputs.
- Accurately received meaningful responses from the Gemini model.
- System handled exceptions like unknown speech or API request errors gracefully.



Learning:

- Understood how to work with microphone input and real-time voice processing in Python.
- Gained hands-on experience with integrating speech recognition with a generative AI model.
- Learned how to structure user interactions in a natural and interactive format using voice commands