

(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.

```

PS C:\Users\jaspi\OneDrive\Desktop\Speech> python app.py
Choose recording mode:
1. Record for fixed duration (e.g., 5 seconds)
2. Record until stop (Ctrl+C or silence)
Enter 1 or 2: 1
Enter duration in seconds (default 5): 4
🔊 Recording for 4 seconds...
📁 Audio saved as: output.wav
🗣️ Converting speech to text...
🗣️ You said: what is generative AI

🤖 Gemini is thinking...

💬 Gemini's Response:
Generative AI refers to a type of artificial intelligence that can create new content. This content can be in the form of:

* Text: Writing articles, poems, code, scripts, etc.
* Images: Generating realistic or stylized images from text descriptions or other input.
* Audio: Creating music, speech, and other sounds.
* Video: Producing video clips or entire movies.
* 3D models: Designing virtual objects and environments.
* Code: Writing or assisting in the generation of computer code.

How it works (in a nutshell):

Generative AI models are typically trained on massive datasets of existing content. During training, the

```

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