Summer Training TR-103

Prompt Engineering

Day-7 Report

The 7th day of the training focused on introducing Speech-to-Text technology and its integration into prompt engineering.

Introduction to Speech-to-Text:

Speech-to-Text is a technology that converts spoken language into written text, enhancing prompt engineering by enabling voice-based inputs for AI models. This session implemented a system using 'sounddevice' for audio recording, 'scipy' for file handling, 'ipywidgets' for an interactive interface, and the OpenAI 'whisper-1' model for transcription.

The ultimate goal is to develop effective voice prompts that result in accurate text outputs.

- **Significance of Speech-to-Text in Prompt Engineering:**
    - Enables real-time voice-to-text conversion.
    - Enhances interactivity with AI systems.
    - Supports dynamic prompt generation.
- **Real-World Use Cases:**
    - **Voice Input:** Record and transcribe commands.
    - **Interactive Tools:** Use in real-time applications.
    - **Testing:** Evaluate transcription accuracy.

Role of Libraries and Widgets in Code:

- **sounddevice:** Captures audio data for recording.
- **scipy:** Handles saving recorded audio as files.
- **ipywidgets:** Creates buttons for interactive control.
- **whisper-1 (OpenAI):** Transcribes recorded audio into text using the OpenAI API.

Functionality of the 3 Buttons:

- **Record Fixed Duration:** Starts a set-duration recording.
- **Record with Stop Button:** Allows manual start and stop of recording.
- **Stop Recording:** Halts the recording process.

Accuracy:

- Accuracy, including punctuation, depends on audio quality and model training.
- Paid versions of the OpenAI API offer higher accuracy for punctuation and more.