DAY – 7

SPEECH-TO-TEXT CONVERSION USING WHISPER API

On Day 7, we shifted focus to the integration of audio input with AI, by learning how to develop a Speech-to-Text Conversion System using Python. The objective was to capture user speech through a microphone and convert it accurately into text using the Whisper model, provided via Hugging Face's API.

This project helped us understand the practical application of voice recognition, and how speech can be transcribed dynamically using AI.

DEPENDENCIES USED

We installed and implemented the following Python libraries:

- **sounddevice** – to access the system microphone and record real-time audio.
- **scipy** – for handling and saving audio data.
- **ipywidgets** – for creating an interactive user interface in Jupyter Notebooks.

These tools helped us build a flexible and interactive environment for recording and analyzing voice inputs.

TWO RECORDING MODES IMPLEMENTED

We developed the functionality to handle two different modes of speech recording:

a. Record for a Fixed Duration

- The user sets a predefined duration (e.g., 5 seconds), speaks into the microphone, and the tool automatically stops recording after the time ends.

b. Record Until Stop (8-Second Timeout)

- The system listens for a maximum of 8 seconds or until silence is detected.
- This approach is more natural for open-ended speech and spontaneous inputs.

SPEECH TRANSCRIPTION USING WHISPER MODEL

After recording, the audio was passed to Whisper, a robust speech recognition model developed by OpenAI and integrated using an API key from Hugging Face.

Whisper returned the transcribed text based on the audio input, and we then evaluated:

- How accurately it captured spoken words

- Whether any words were missed or misinterpreted
- The difference in performance across both recording modes

This helped us assess the reliability of AI-based transcription in real-world scenarios.

CONCLUSION

Day 7 introduced us to the world of AI-powered speech recognition, where we coded a working prototype for converting voice to text using the Whisper model. This hands-on session enhanced our understanding of:

- Real-time audio processing
- Interactive UI elements in Python
- API integration with Hugging Face
- Evaluating transcription accuracy

This learning is valuable for developing intelligent applications such as voice assistants, transcription tools, and accessibility support systems.