Guru Nanak Dev Engineering College

Training Diary – TR-102 Report

Name: Divanshi Goyal
URN: 2302513
CRN: 2315056
Day 6

Training Summary

On the sixth day of training, we focused on integrating **speech-based input systems** into our AI projects. We built a **Speech-to-Text Converter** using OpenAI's **Whisper model** and extended that functionality to improve our earlier projects: **URL Summarizer**, **Keyword Searcher**, and **Image Generator**.

Project 1: Speech-to-Text Converter using Whisper

We developed a tool that:

- Records user voice input through a microphone.
- Saves the audio in **WAV format**.
- Uses the **Whisper model** to transcribe speech into text.
- Adds **punctuation** and formatting for cleaner and more readable output.

This provided us with a robust way to turn natural speech into usable text for downstream applications.

Project 2: Upgrading URL Summarizer, Keyword Searcher & Image Generator

After successfully building the speech-to-text component, we integrated it into the following AI tools:

🔗 **URL Summarizer (Enhanced)**

- Now supports **voice input** where the user can speak a website URL.
- The spoken URL is converted to text using Whisper and then summarized as before.

🔍 **Keyword Searcher (Enhanced)**

- Users can now **speak keywords** instead of typing them.
- The model checks if the keyword is found in the summary or raw content of the URL.

- Returns contextually relevant information or states that the keyword is not found.

🖼️ **Image Generator (Enhanced)**

- Accepts **voice prompts** to generate images.
- The prompt is transcribed using Whisper and passed to Hugging Face's image generation API.
- Metadata and prompt history are stored in a JSON file, as before.

This integration allowed us to make our tools **more accessible** and **hands-free**, showcasing how **voice interfaces** can enhance the usability of AI applications.

Learning Outcome

From today's session, we learned:

- How to build and use Whisper for real-time voice transcription.
- How to handle **audio data as input** for AI-driven tools.
- How to make existing projects more **user-friendly** and **multi-modal** by combining audio and AI.