Guru Nanak Dev Engineering College

Training Diary – TR-102 Report

Name: Dilnaz Kaur Grewal
URN: 2302510
CRN: 2315054
Day 7

Training Summary

On the seventh day of training, we explored the complete cycle of voice-based AI systems by building:

- A **Text-to-Speech (TTS)** converter using **xtts**, and
- A **Speech-to-Speech (S2S)** converter by combining **Whisper** and **xtts**.

These applications allow machines to understand human speech and respond audibly, forming the foundation for voice assistants, dubbing systems, and accessibility tools.

Project 1: Text-to-Speech Converter (TTS)

We first created a Text-to-Speech tool using **xtts (Extended Text-to-Speech)**, which converts written input into spoken audio.

Steps Involved:

- Accept **text input** from the user.
- Process the text using **xtts**, a neural network trained for lifelike speech synthesis.
- Generate and play the **speech output**.
- Save the audio as a file (e.g., .wav format) for playback or storage.

Project 2: Speech-to-Speech Converter (S2S)

We then combined two components — **Speech-to-Text (STT)** and **Text-to-Speech (TTS)** — to build a full **Speech-to-Speech Converter**.

🔁 How It Works – Speech-to-Speech Architecture

1. **Speech-to-Text (STT) – Using Whisper**
    - User speaks into a microphone.
    - The speech is captured and saved as an audio file (usually .wav).

- **Whisper**, an automatic speech recognition model by OpenAI, transcribes the voice into **clean and punctuated text**.

2. **Text-to-Speech (TTS) – Using xtts**
    - The transcribed text is passed as input to the **xtts model**.
    - xtts generates a **new audio file** from the text in a synthetic voice.
    - The final audio is played and saved.

🛠 Pipeline Summary

Speech Input → Whisper (STT) → Text → xtts (TTS) → Speech Output

This full-cycle conversion mimics natural conversation and can be used in real-world systems like:

- Multilingual AI voice translators
- Assistive communication devices
- Interactive AI storytelling
- Custom voicebots and dubbing tools

Learning Outcome

By completing this integrated system, we learned:

- How to chain multiple AI models to build a voice-based application.
- How to manage **audio input/output**, transcriptions, and synthesized responses.
- Real-world use cases for **voice-to-voice automation**.