DAY – 8

TEXT-TO-SPEECH & SPEECH-TO-SPEECH CONVERSION USING AI

On Day 8 of our training, we explored advanced audio transformation techniques using Artificial Intelligence. The session focused on converting text into speech and further extending that capability to enable full speech-to-speech generation using the Eleven Labs API, a popular platform for high-quality AI-generated voices.

1. TEXT-TO-SPEECH CONVERSION

We began by developing a Python-based Text-to-Speech (TTS) system. The process included the following steps:

- The user inputs any text string dynamically.
- The system sends this text to Eleven Labs via a valid API key.
- The AI then returns a high-quality audio output of the given text in a human-like voice.

This allowed us to:

- Experiment with different types of text inputs (e.g., formal, casual, questions).
- Hear the voice output generated in real time.
- Understand the capabilities of Eleven Labs in generating natural-sounding speech.

This task has practical use cases in areas like voice assistants, audiobook generation, accessibility tools, and more.

2. SPEECH-TO-SPEECH GENERATION

We then moved on to a more comprehensive task — Speech-to-Speech conversion, which integrates multiple steps:

a. Voice Input

- The user speaks into the microphone using a real-time audio recorder.
- The audio is captured and converted to text using speech recognition.

b. Text Processing

- The spoken input is shown on the terminal as text.
- The text is processed (can include sentiment adjustment, language translation, or summarization depending on the project scope).

c. Output Generation

- The processed text is converted back to speech using Eleven Labs.
- The AI-generated audio response is then played in real time.

d. File Saving

- The final output audio is saved automatically in both .wav and .mp3 formats to the user's system.
- This allows for reuse or integration into other applications.

This task helped us understand how we can build complete conversational interfaces using AI tools. It also showed how multiple models and APIs can be chained together to build seamless human-computer voice interactions.

TOOLS & TECHNOLOGIES USED

- **Eleven Labs API** – for generating lifelike speech from text.
- **Speech Recognition Library** – to transcribe audio input to text.
- **Python Sound Libraries** – for recording, playback, and saving audio in different formats.

CONCLUSION

Day 8 was a highly interactive and technical session focused on voice technologies powered by AI. We successfully developed:

- A fully functional Text-to-Speech system.
- A complete Speech-to-Speech pipeline involving real-time voice capture, transcription, AI processing, and audio response.

These skills have direct applications in the development of AI voice assistants, customer support bots, educational tools, and accessibility features for users with special needs.