# Training Day 7 Report:

**Speech-to-Text Conversion using LLMs like Gemini**

**Speech-to-text conversion** is the process of converting spoken audio into written text using artificial intelligence. With the help of **large language models (LLMs)** like **Google Gemini**, this task can now be done with high accuracy, even in noisy environments or with different accents.

- **Step 1 – Input Audio:**
  The first step is recording or uploading a speech file. This could be a voice note, podcast, phone call, or live microphone input.
- **Step 2 – Model Selection:**
  Tools like **Google Gemini**, **Whisper**, or other AI APIs are selected to process the audio. These models are trained on large datasets of spoken language and can understand different languages, dialects, and tones.
- **Step 3 – Audio Preprocessing:**
  Before sending to the model, the audio is usually cleaned by removing background noise and converting it to the correct format (like .wav or .mp3). This ensures better accuracy during transcription.
- **Step 4 – Transcription by LLM:**
  The audio is sent to the AI model. It listens to the sound and uses its training to predict the corresponding text. In multimodal models like **Gemini**, this step can also be combined with text or visual inputs for deeper understanding (e.g., transcribing and summarizing).
- **Step 5 – Output and Usage:**
  The final output is clean, readable text. This can be used in meeting notes, subtitles, accessibility tools, virtual assistants, and more.
- **Tools Used:**
  - **Google Gemini / Whisper** – To convert audio to text using AI.
  - **Python Libraries (SpeechRecognition, pydub)** – For handling and processing audio.

  - **Microphone or Audio Files** – As the source of spoken input.

- **Why It's Useful:**
  Speech-to-text helps people interact with computers through voice. It supports fast note-taking, helps those with hearing impairments, powers smart assistants, and saves time in writing or transcription work.

In summary, **speech-to-text conversion** using modern AI tools like **Google Gemini** is accurate, fast, and adaptable. It makes it easier to bridge the gap between human speech and digital content, enhancing productivity and accessibility.