

Summer Training TR-103 Prompt Engineering

Day 8 Report

Text-to-Speech

Today, I explored how to convert user input in text format into audible speech using the Gemini API. The process involved taking a written prompt, generating a relevant response using the Gemini model, and converting that response into spoken output using an offline text-to-speech engine. This feature enhances the overall user experience by providing interactive and accessible voice responses. It is particularly useful in applications like virtual assistants, educational tools, and accessibility-focused platforms. The seamless transition from text to speech demonstrated how effectively AI can bridge communication between humans and machines.

Speech-to-Speech

I also worked on building a complete voice-based assistant that supports real-time spoken interactions. The system was designed to capture voice input through a microphone, convert the speech into text using speech recognition, generate an intelligent reply using the Gemini API, and then convert the AI-generated text back into speech. In addition to real-time interaction, the system could log conversations and save audio responses for future use. This practical implementation showed how different technologies can be integrated to create efficient, user-friendly AI assistants that communicate naturally and respond instantly.

Music Generation

Learned how to generate background music using a text-to-audio model. By providing a descriptive music prompt (e.g., “an uplifting and gentle melody with soft piano and light orchestral strings”), the model successfully created a corresponding audio file. The output was saved as a .wav file for playback or integration into multimedia projects. This functionality is

especially useful for enhancing user experiences in games, apps, podcasts, and creative media without needing manual music composition.