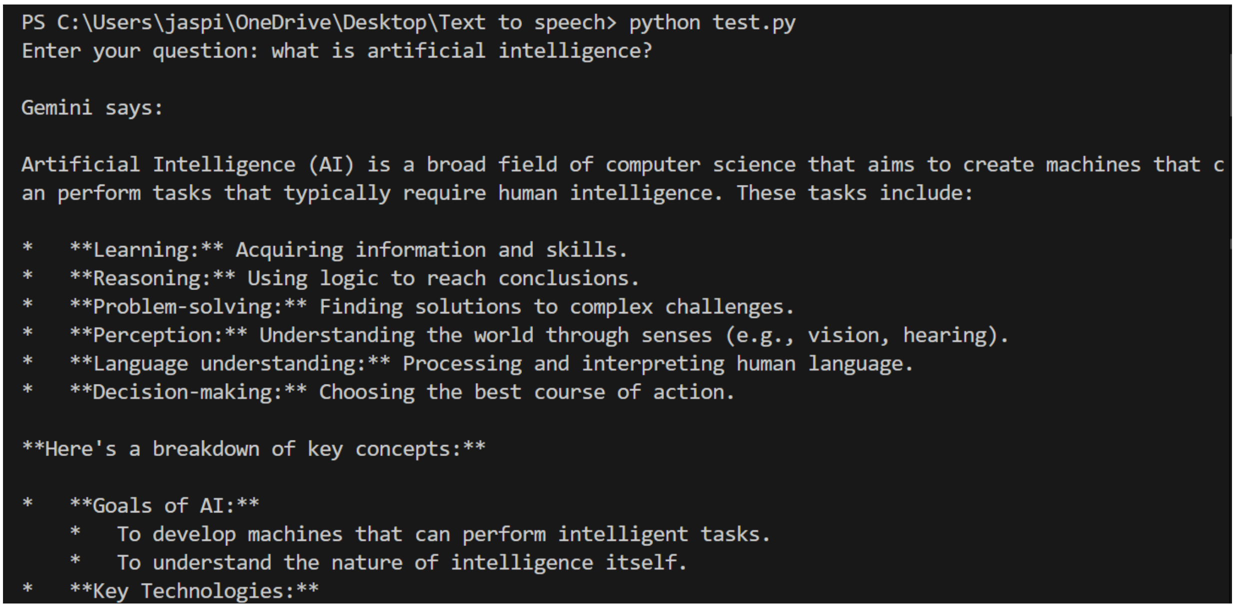
(TR-103) PROMPT ENGINEERING –

Training Day 8 Report:

Text-to-Speech, Speech-to-Speech and Music Generation using Gemini API

Text-to-Speech:

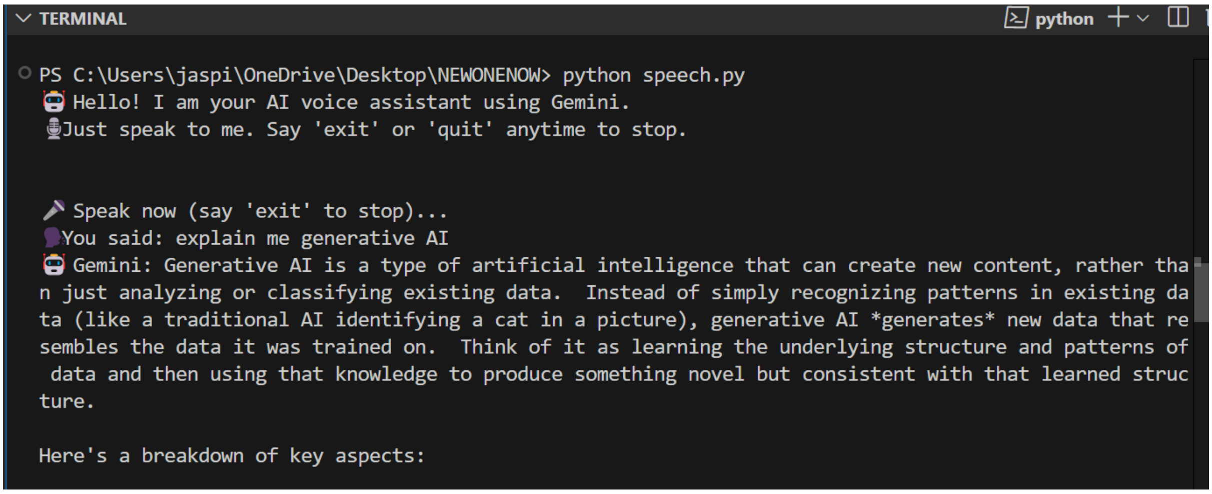
Today, I explored how to convert user input in text format into audible speech using the Gemini API. The process involved taking a written prompt, generating a relevant response using the Gemini model, and converting that response into spoken output using an offline text-to-speech engine. This feature 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.



```
PS C:\Users\jaspi\OneDrive\Desktop\Text to speech> python test.py
Enter your question: what is artificial intelligence?

Gemini says:

Artificial Intelligence (AI) is a broad field of computer science that aims to create machines that c
an perform tasks that typically require human intelligence. These tasks include:

*   **Learning:** Acquiring information and skills.
*   **Reasoning:** Using logic to reach conclusions.
*   **Problem-solving:** Finding solutions to complex challenges.
*   **Perception:** Understanding the world through senses (e.g., vision, hearing).
*   **Language understanding:** Processing and interpreting human language.
*   **Decision-making:** Choosing the best course of action.

**Here's a breakdown of key concepts:**

*   **Goals of AI:**
    *   To develop machines that can perform intelligent tasks.
    *   To understand the nature of intelligence itself.
*   **Key Technologies:**
```

Speech-to-Speech:

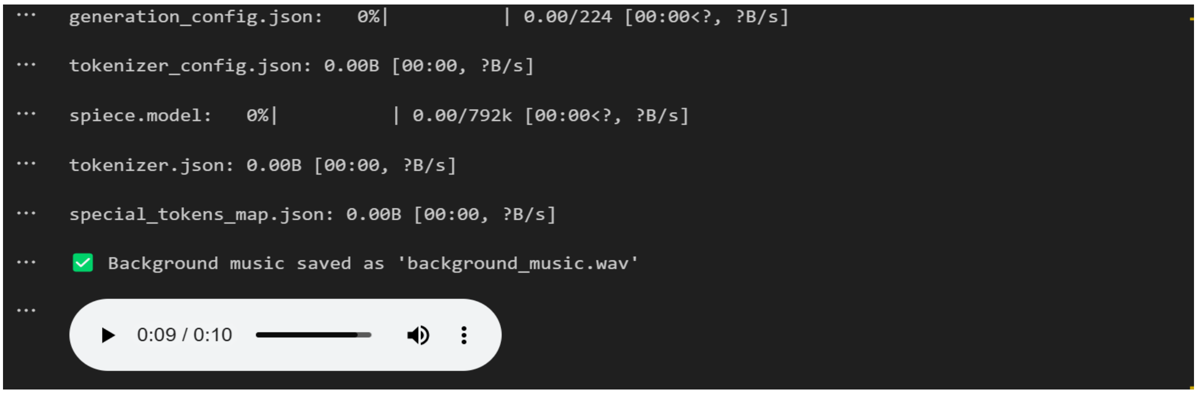
I also worked on building a complete voice-based assistant that supports real-time spoken interactions. The system was designed to capture voice input through a microphone, convert the speech into text using speech recognition, generate an intelligent reply using the Gemini API, and then convert the 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.



```
○ PS C:\Users\jaspi\OneDrive\Desktop\NEWONENOW> python speech.py
  🤖 Hello! I am your AI voice assistant using Gemini.
  🎙Just speak to me. Say 'exit' or 'quit' anytime to stop.


  🎤 Speak now (say 'exit' to stop)...
  🗣You said: explain me generative AI
  🤖 Gemini: Generative AI is a type of artificial intelligence that can create new content, rather tha
  n just analyzing or classifying existing data.  Instead of simply recognizing patterns in existing da
  ta (like a traditional AI identifying a cat in a picture), generative AI *generates* new data that re
  sembles the data it was trained on.  Think of it as learning the underlying structure and patterns of
   data and then using that knowledge to produce something novel but consistent with that learned struc
  ture.

  Here's a breakdown of key aspects:
```

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.



```
…   generation_config.json:   0%|          | 0.00/224 [00:00<?, ?B/s]

…   tokenizer_config.json: 0.00B [00:00, ?B/s]

…   spiece.model:   0%|          | 0.00/792k [00:00<?, ?B/s]

…   tokenizer.json: 0.00B [00:00, ?B/s]

…   special_tokens_map.json: 0.00B [00:00, ?B/s]

…   ✅ Background music saved as 'background_music.wav'

…
```