Real Time Speech to Text to Speech

Leveraging OpenAI for Transcription, Response Generation, and Speech Synthesis

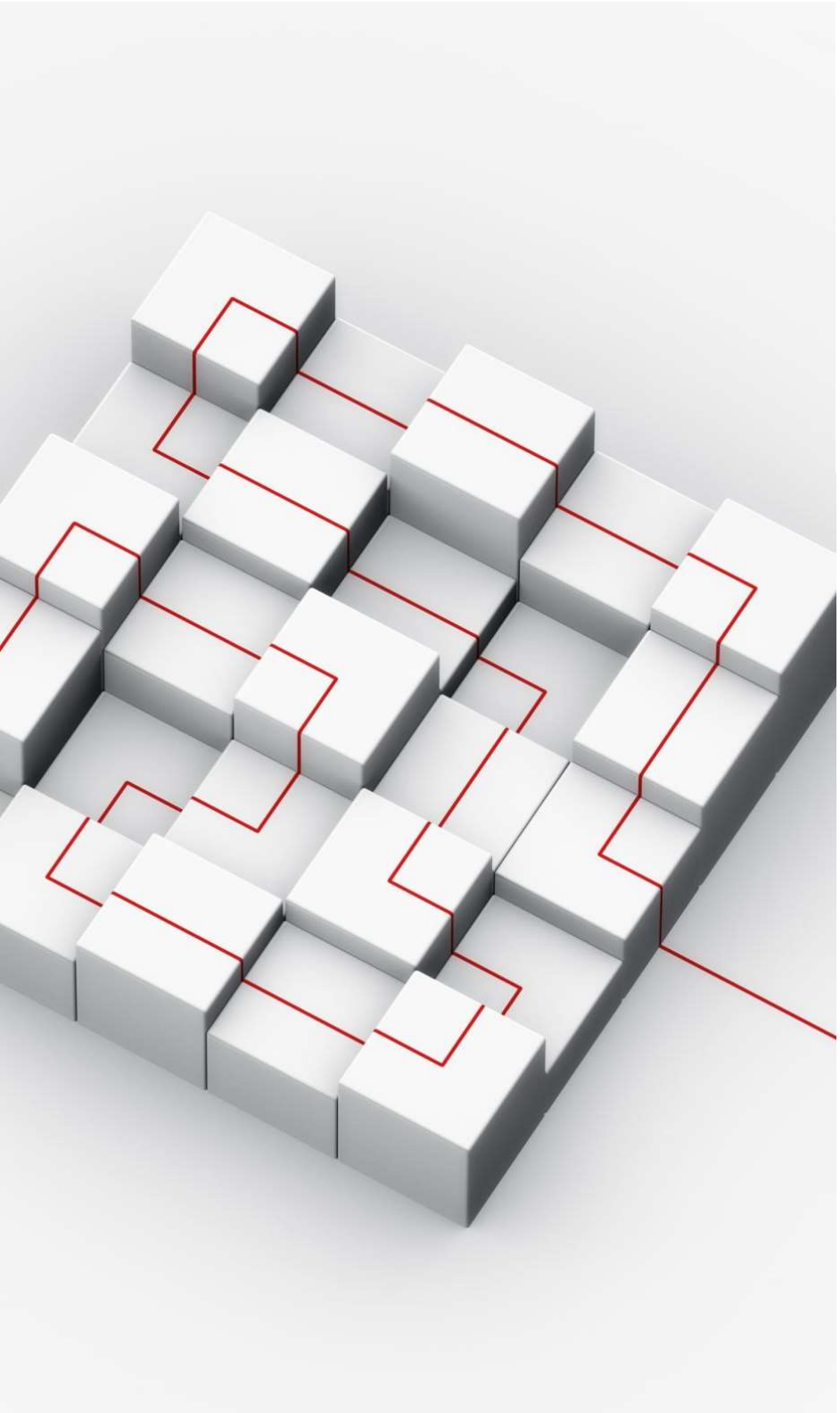
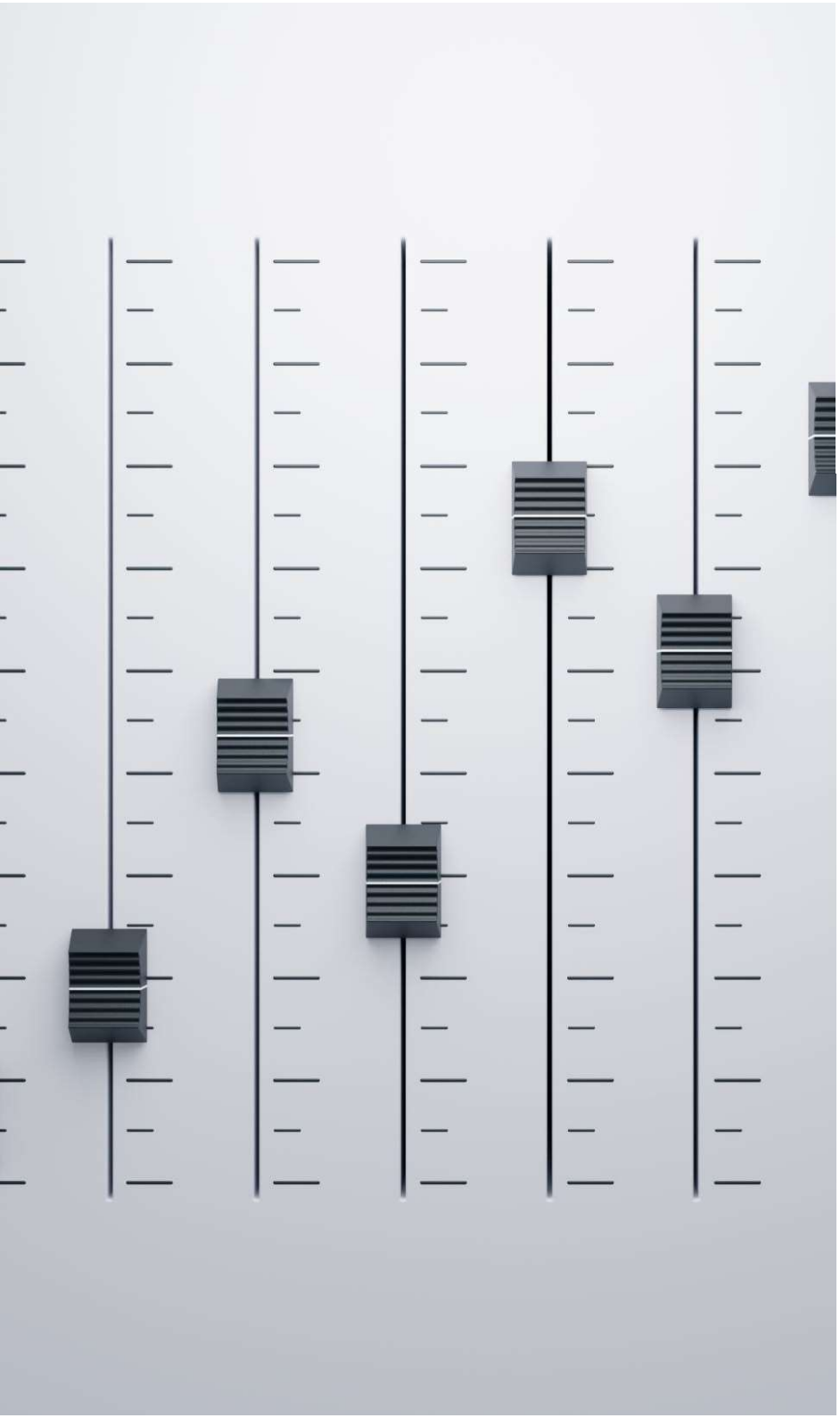


Table of Contents

Introduction

This project utilizes OpenAI's Whisper for speech-to-text, GPT-3.5 for response generation, and a hypothetical OpenAI text-to-speech (TTS) model.

The goal is to process an audio file, generate a text-based response using GPT, and convert that response to speech.



Design

The design revolves around three main steps:

1. Audio Transcription using OpenAI Whisper model.
2. Text Generation with GPT-3.5.
3. Speech Synthesis using OpenAI TTS (hypothetical model).

- The choice of OpenAI models is aimed at ensuring high-quality results in each stage of the process.

Why This Approach?

We chose OpenAI's Whisper, GPT-3.5, and a TTS system because of their cutting-edge performance and versatility in natural language processing.

The goal was to use a seamless, integrated solution for transcribing, responding, and synthesizing speech.

Investigating Solutions

Other potential solutions considered were:

1. Google Cloud Speech-to-Text API for transcription.
2. Using other TTS services like Amazon Polly.

However, OpenAI's Whisper and GPT-3.5 were chosen for their superior accuracy in transcription and generation.

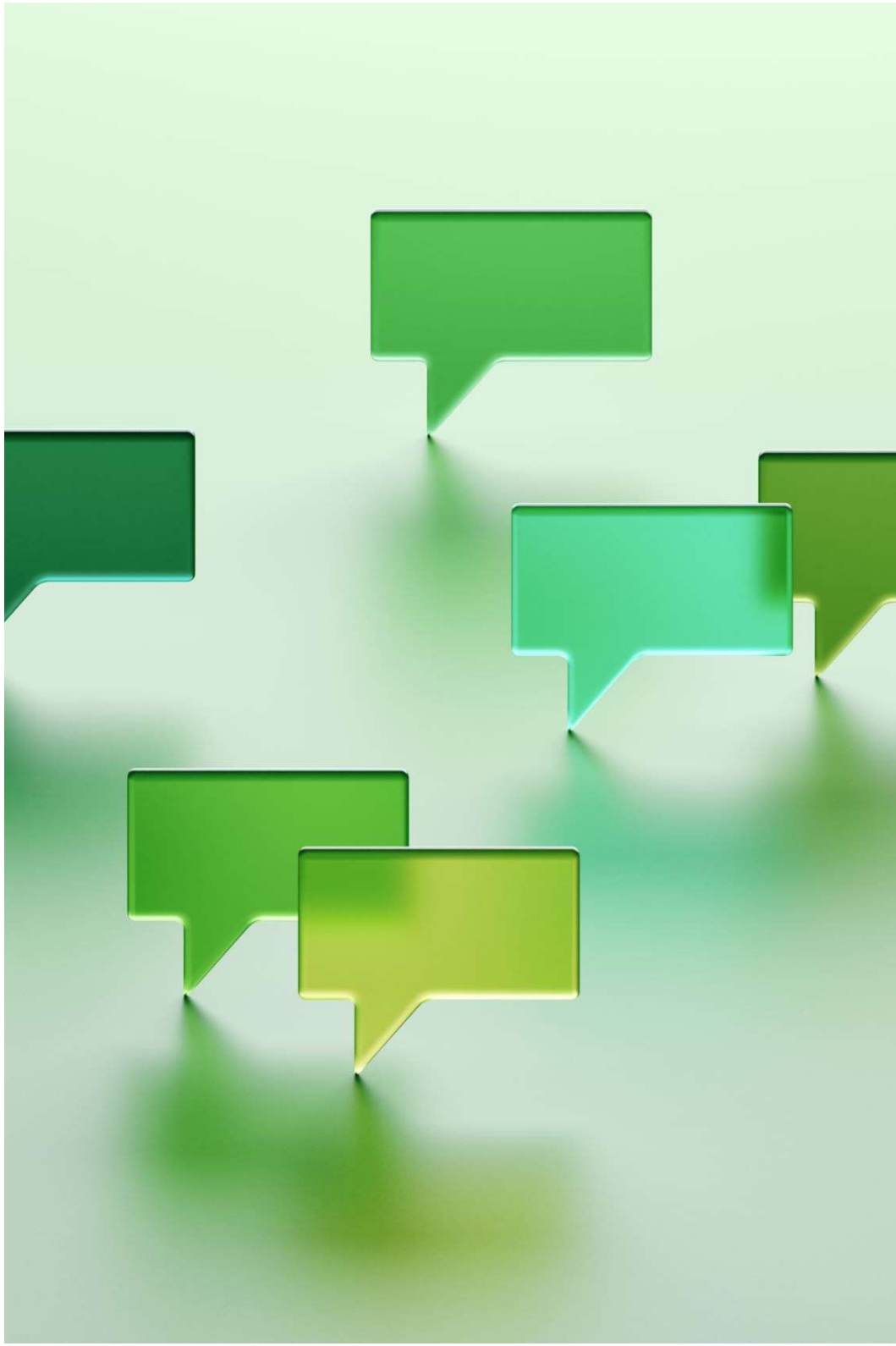
Implementation

The implementation process includes the following steps:

1. Transcribing Audio: The audio is uploaded to OpenAI's Whisper model for transcription.

2. Generating GPT Response: The transcribed text is sent to GPT-3.5 to generate a relevant response.

3. Text-to-Speech: The GPT response is passed to a TTS model to synthesize speech output.



Test

Testing was conducted on various audio files, ensuring:

1. Transcription Accuracy: Accurate conversion of speech to text using Whisper.
2. Response Quality: Relevant and coherent responses generated by GPT-3.5.
3. Speech Clarity: Clear and natural-sounding speech synthesis from the GPT-generated text.

Enhancement Ideas

Future ideas for improvement include:

1. Real-Time Transcription and Response: Enable live audio input processing.
2. Multilingual Support: Support transcription and speech synthesis in multiple languages.
3. Customizable Voice: Offer options to select or customize the voice for TTS.

Conclusion

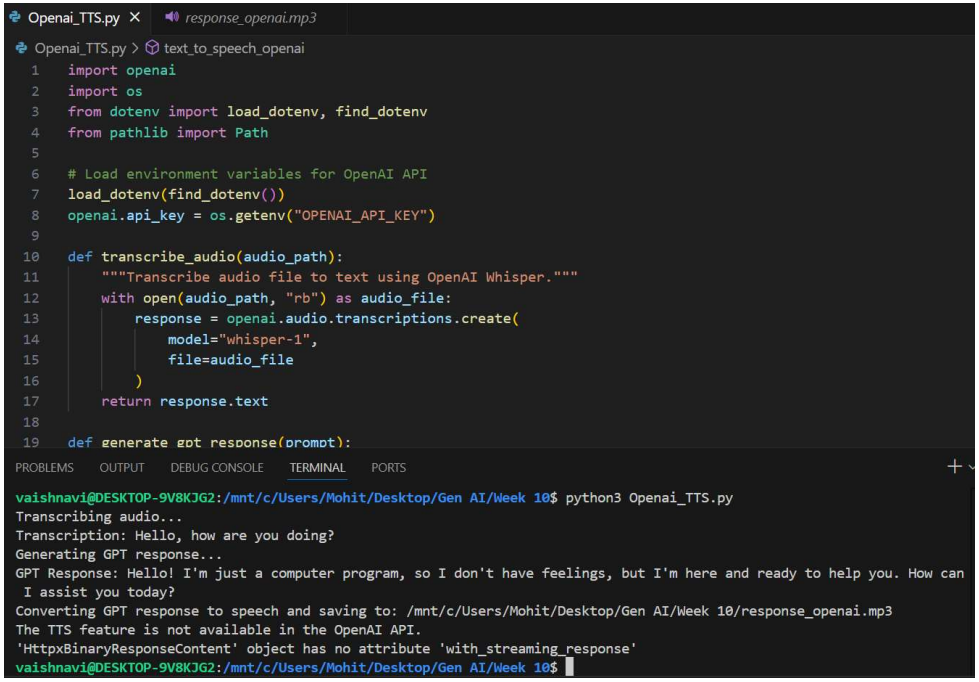
This project successfully integrates OpenAI's models for creating a voice interaction system that transcribes speech, generates responses, and converts them to speech.

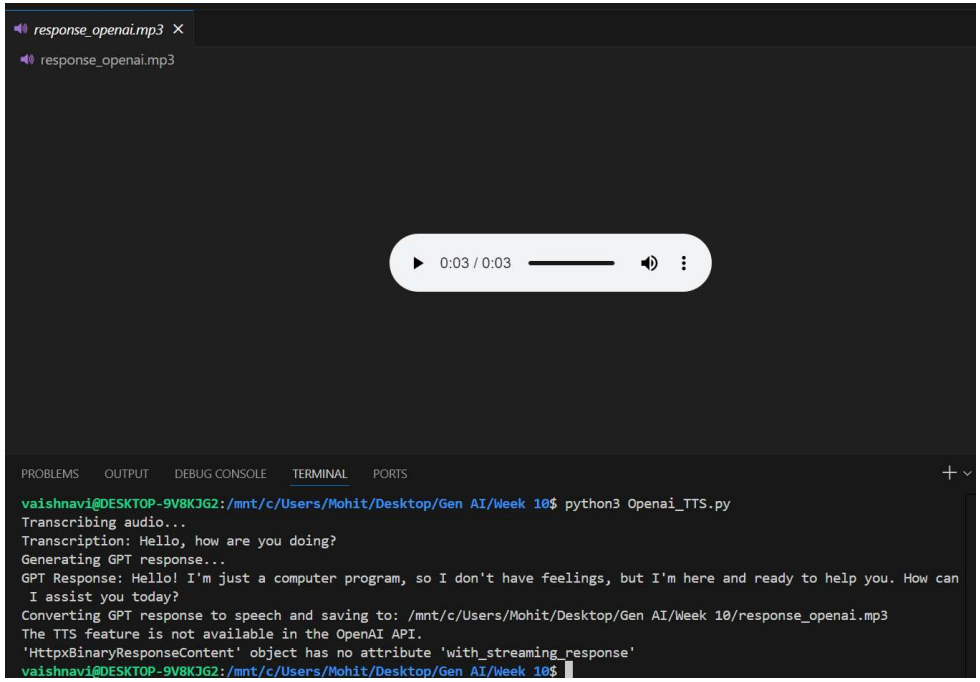
The system is highly modular, allowing for future improvements.

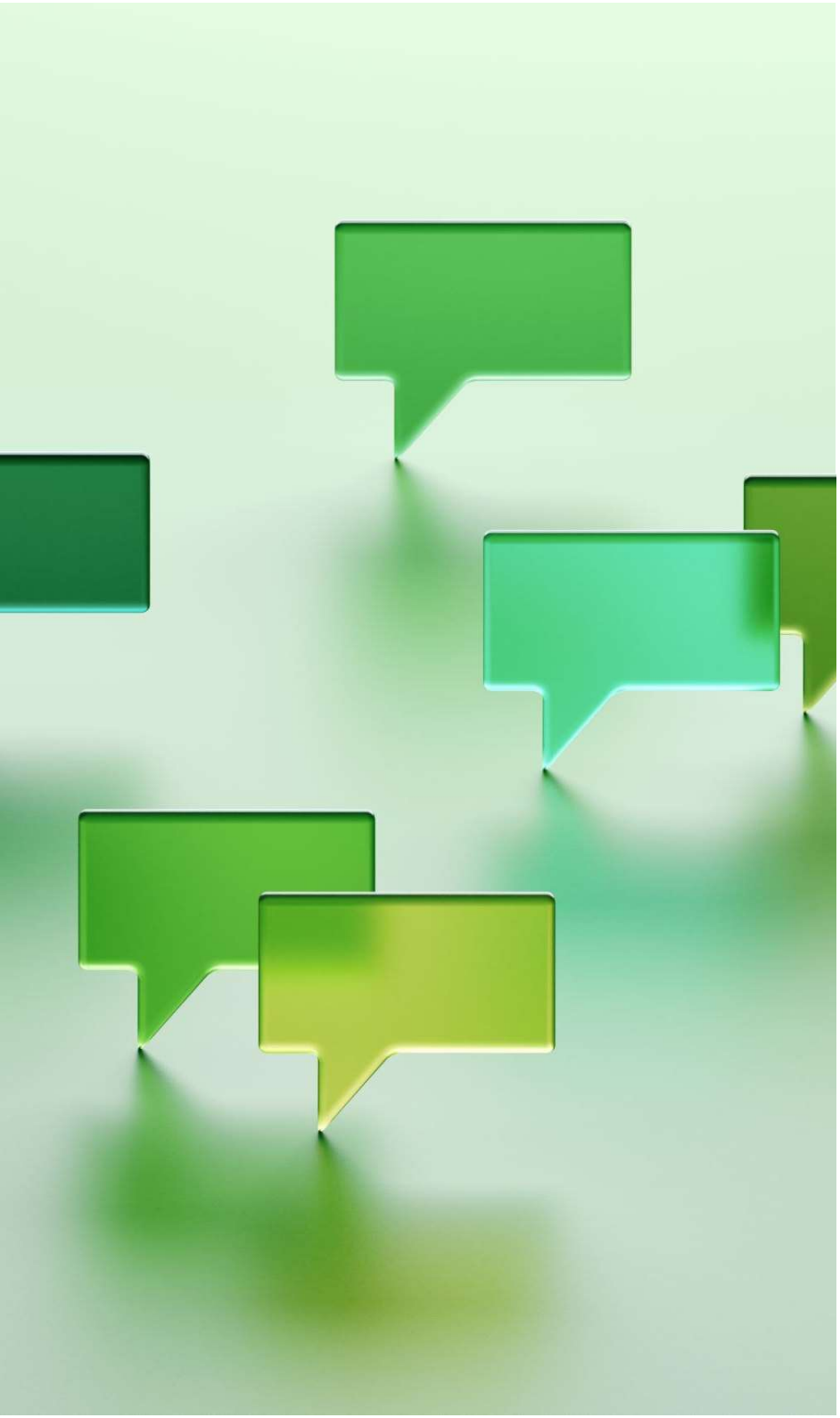
References

- 1. OpenAI Whisper:https://openai.com/research/whisper
- 2. OpenAI GPT-3.5:https://openai.com/research/gpt-3
- 3. OpenAI API Documentation: https://platform.openai.com/docs
- 4. Python Requests:https://requests.readthedocs.io/en/master/

Appendix


Openai_TTS.py
response_openai.mp3
Openai_TTS.py > text_to_speech_openai
1 import openai
2 import os
3 from dotenv import load_dotenv, find_dotenv
4 from pathlib import Path
5
6 # Load environment variables for OpenAI API
7 load_dotenv(find_dotenv())
8 openai.api_key = os.getenv("OPENAI_API_KEY")
9
10 def transcribe_audio(audio_path):
11 """Transcribe audio file to text using OpenAI Whisper."""
12 with open(audio_path, "rb") as audio_file:
13 response = openai.audio.transcriptions.create(
14 model="whisper-1",
15 file=audio_file
16)
17 return response.text
18
19 def generate_gpt_response(prompt):
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS
vaishnavi@DESKTOP-9V8KJG2:/mnt/c/Users/Mohit/Desktop/Gen AI/Week 10$ python3 Openai_TTS.py
Transcribing audio...
Transcription: Hello, how are you doing?
Generating GPT response...
GPT Response: Hello! I'm just a computer program, so I don't have feelings, but I'm here and ready to help you. How can I assist you today?
Converting GPT response to speech and saving to: /mnt/c/Users/Mohit/Desktop/Gen AI/Week 10/response_openai.mp3
The TTS feature is not available in the OpenAI API.
'HttpxBinaryResponseContent' object has no attribute 'with_streaming_response'
vaishnavi@DESKTOP-9V8KJG2:/mnt/c/Users/Mohit/Desktop/Gen AI/Week 10$


response_openai.mp3
response_openai.mp3
0:03 / 0:03
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS
vaishnavi@DESKTOP-9V8KJG2:/mnt/c/Users/Mohit/Desktop/Gen AI/Week 10$ python3 Openai_TTS.py
Transcribing audio...
Transcription: Hello, how are you doing?
Generating GPT response...
GPT Response: Hello! I'm just a computer program, so I don't have feelings, but I'm here and ready to help you. How can I assist you today?
Converting GPT response to speech and saving to: /mnt/c/Users/Mohit/Desktop/Gen AI/Week 10/response_openai.mp3
The TTS feature is not available in the OpenAI API.
'HttpxBinaryResponseContent' object has no attribute 'with_streaming_response'
vaishnavi@DESKTOP-9V8KJG2:/mnt/c/Users/Mohit/Desktop/Gen AI/Week 10$



- **Google Slide:**
https://docs.google.com/presentation/d/1DKG3c81CK5AYlQfIz0x2s1xXyQcpPLp6EVsV9rNY3O4/edit?usp=sharing

- **GitHub URL:**
https://github.com/vaishnavi477/Machine-Learning/tree/main/AI-Based Alexa/Real-time Speech-to-Text-to-Speech/OpenAI TTS

Thank You