Real Time Speech to Text to Speech

Leveraging OpenAI and gTTS for Voice-based Interaction

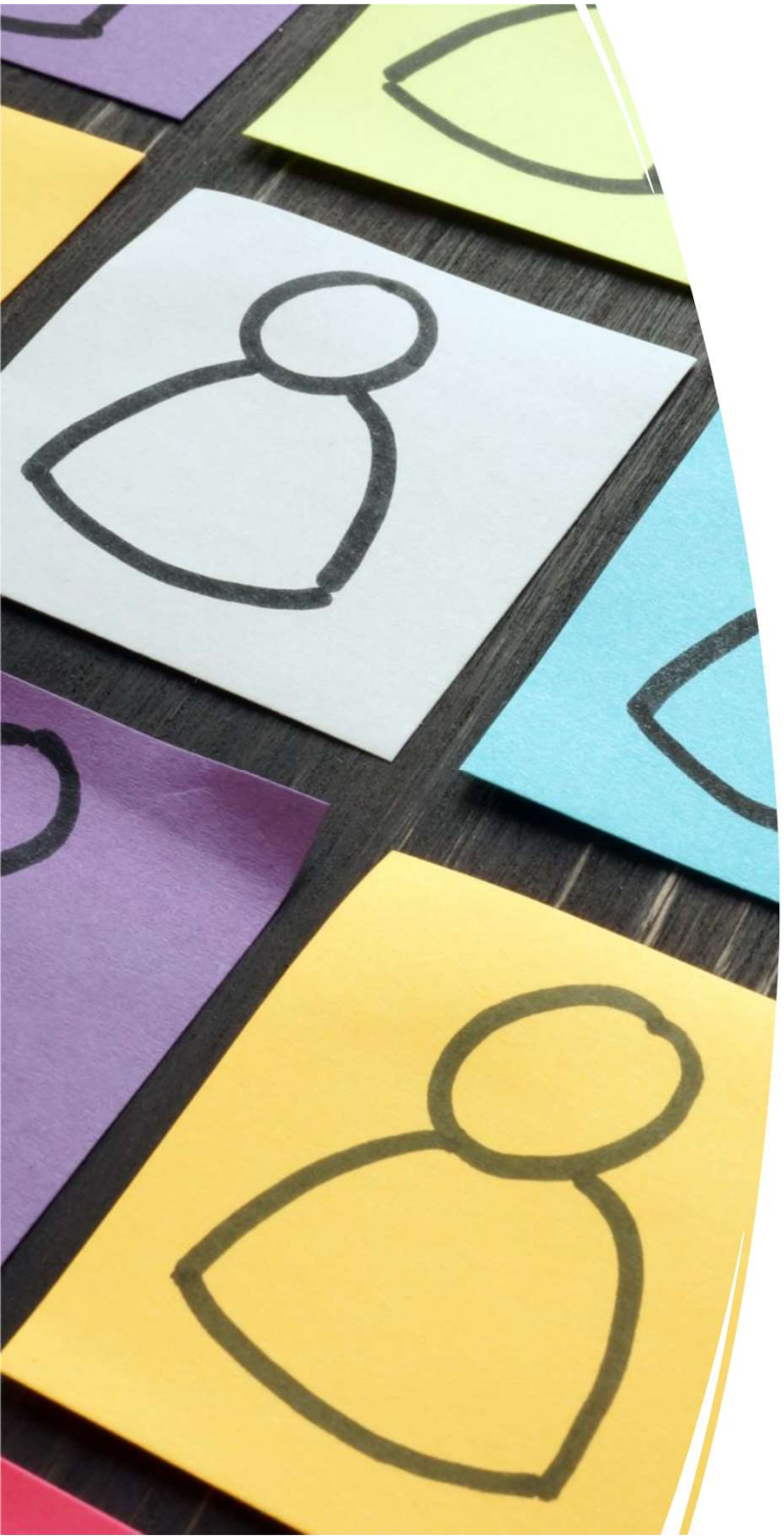


Table of Contents

Introduction

This project leverages OpenAI's advanced technologies, including the Whisper model for speech-to-text transcription, GPT-3.5 for conversational AI, and Google Text-to-Speech (gTTS) for converting text back into speech.

The core functionality of the project revolves around transcribing an audio file, generating a response based on the transcription, and finally converting the response into an audio output.

Design

The goal of the system is to transcribe an audio file, generate a response, and convert that response back into an audio format. Here's the thought process behind the design:

Identify and understand the problems:

1. Speech Recognition
2. Natural Language Processing
3. Speech Synthesis

Investigate possible solutions:

1. Whisper for transcription
2. GPT-3.5 for response generation
3. gTTS for speech synthesis

Theoretical comparison and selection of the best solution:

Whisper for transcription, GPT-3.5 for response generation, and gTTS for speech synthesis.

Implementation

- The implementation involves three main stages:

1. Transcribing Audio to Text: The audio file is loaded, and the Whisper model transcribes it.
2. Generating a Response Using GPT: The transcribed text is used as input for the GPT-3.5 model to generate a response.
3. Converting Text to Speech: The response from GPT-3.5 is converted into speech using gTTS.

- Each of these steps is executed sequentially in the main function, ensuring a smooth flow.

Test

- The system was tested with various audio inputs, focusing on:
    1. Accuracy of speech-to-text transcription
    2. Relevance of GPT responses
    3. Quality of speech output

Sample Test Case:

Input: Hello, How are you doing?'

Transcription: ‘Hello, How are you doing?'

GPT Response: ‘Hello! I’m here and ready to help you. How can I assist you today?'

Speech Output: Clear speech saved as an MP3 file.

Enhancement Ideas

Enhancement ideas include:

1. Real-time Transcription and Response
2. Multiple Language Support
3. Voice Activity Detection
4. Custom Voice Output
5. Error Handling

Conclusion

The project successfully integrates Whisper for transcription, GPT-3.5 for response generation, and gTTS for speech synthesis.

The system demonstrates how combining these technologies can create an interactive voice-based system, making it efficient and intuitive for users.

References

1. OpenAI API Documentation: https://platform.openai.com/docs

2. Whisper:https://openai.com/research/whisper

3. GPT-3.5:https://openai.com/research/gpt-3

4. gTTS (Google Text-to-Speech): https://pypi.org/project/gTTS/

5. Python Documentation: https://docs.python.org/3/

Appendix

response_gtts.mp3 | Google_TTS.py 1

Google_TTS.py > ...

```python
import openai
import os
from dotenv import load_dotenv, find_dotenv
from gtts import gTTS
from pathlib import Path

# Load environment variables for OpenAI API
load_dotenv(find_dotenv())
openai.api_key = os.getenv("OPENAI_API_KEY")

def transcribe_audio_to_text(audio_file_path):
    """Transcribe audio file to text using OpenAI Whisper."""
    with open(audio_file_path, "rb") as audio_file:
        transcription = openai.audio.transcriptions.create(
            model="whisper-1",
            file=audio_file
        )
    return transcription.text

```

PROBLEMS 1 OUTPUT DEBUG CONSOLE TERMINAL PORTS

```
vaishnavi@DESKTOP-9V8KJG2:/mnt/c/Users/Mohit/Desktop/Gen AI/Week 10$ python3 Google_TTS.py
Transcribing audio from: audio.m4a
Transcription: Hello, how are you doing?
Generating GPT response...
GPT Reply: 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 reply to speech and saving to: /mnt/c/Users/Mohit/Desktop/Gen AI/Week 10/response_gtts.mp3
Process completed successfully!
vaishnavi@DESKTOP-9V8KJG2:/mnt/c/Users/Mohit/Desktop/Gen AI/Week 10$
```

response_gtts.mp3 | Google_TTS.py 1

response_gtts.mp3

0:00 / 0:09

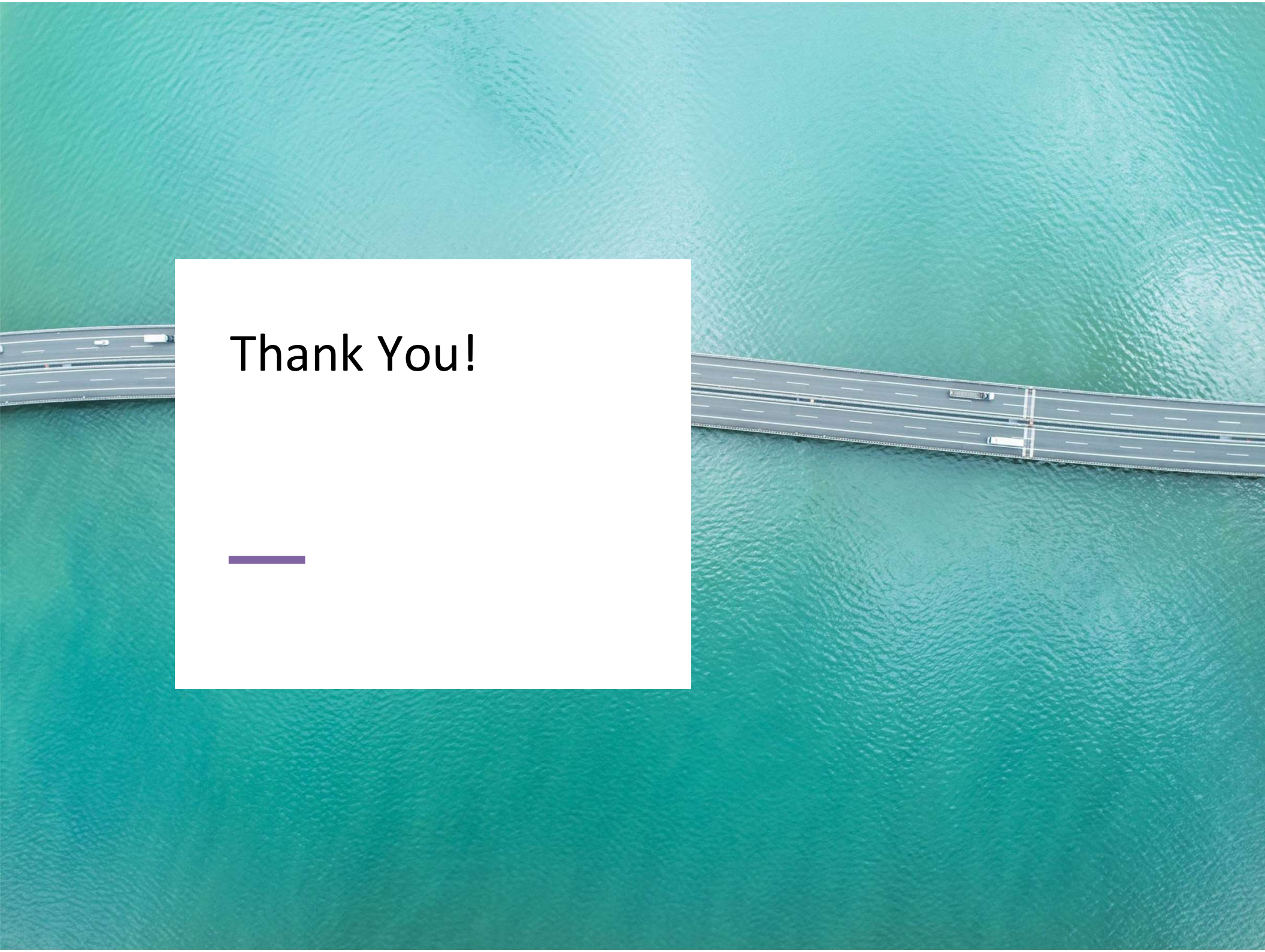
PROBLEMS 1 OUTPUT DEBUG CONSOLE TERMINAL PORTS

```
vaishnavi@DESKTOP-9V8KJG2:/mnt/c/Users/Mohit/Desktop/Gen AI/Week 10$ python3 Google_TTS.py
Transcribing audio from: audio.m4a
Transcription: Hello, how are you doing?
Generating GPT response...
GPT Reply: 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 reply to speech and saving to: /mnt/c/Users/Mohit/Desktop/Gen AI/Week 10/response_gtts.mp3
Process completed successfully!
vaishnavi@DESKTOP-9V8KJG2:/mnt/c/Users/Mohit/Desktop/Gen AI/Week 10$
```

Google Slide:
https://docs.google.com/presentation/d/1TfLGVumZZ-kmxRvvqp8r3711VuyUKhVgFO1C3qqzybI/edit#slide=id.p1

GitHub URL:

https://github.com/vaishnavi477/Machine-Learning/upload/main/AI-Based%20Alexa/Real-time%20Speech-to-Text-to-Speech/Google%20TTS


Thank You!