DS565 - Generative AI-Driven Intelligent Apps Development

Project: Speech to text-text to speech(Google)

Course: DS565
Professor: Dr. Henry Chang

Presented by: Karan Shrestha (20087)

Introduction to OpenAI API

- **What is OpenAI API?**
  - Provides access to powerful language models (GPT series)
  - Applications: Text generation, chatbots, translation, etc.
- **Why GPT-4o Mini?**
  - Efficient, lightweight version of GPT-4, optimized for quick interactions

Flask-based web application for real-time voice-driven AI interaction

Features:

- **Audio transcription using Whisper**
- **Response generation with GPT-3.5 Turbo**
- **Text-to-Speech feedback using Google TTS**

Core Technologies

- **Flask**: Backend framework
- **Whisper**: Audio transcription model
- **OpenAI GPT-3.5 Turbo**: AI response generation
- **gTTS**: Text-to-Speech conversion
- **Werkzeug**: Secure file handling
- **Dotenv**: API key management

System Architecture

1. **Audio Upload**: Users upload their audio files via the web interface.
2. **Transcription**: Audio is transcribed using Whisper.
3. **Response Generation**: GPT-3.5 generates an intelligent response.
4. **TTS Feedback**: The response is converted into speech and returned to the user.

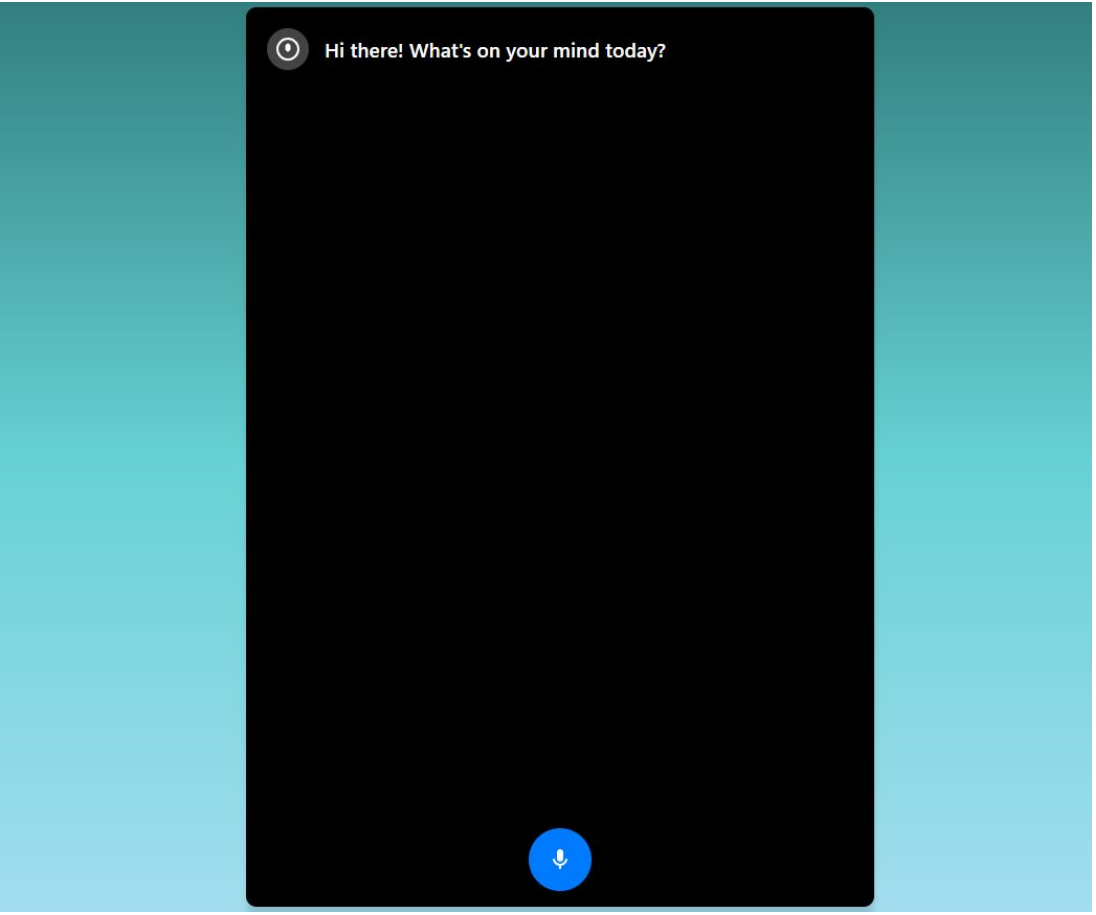
Highlights

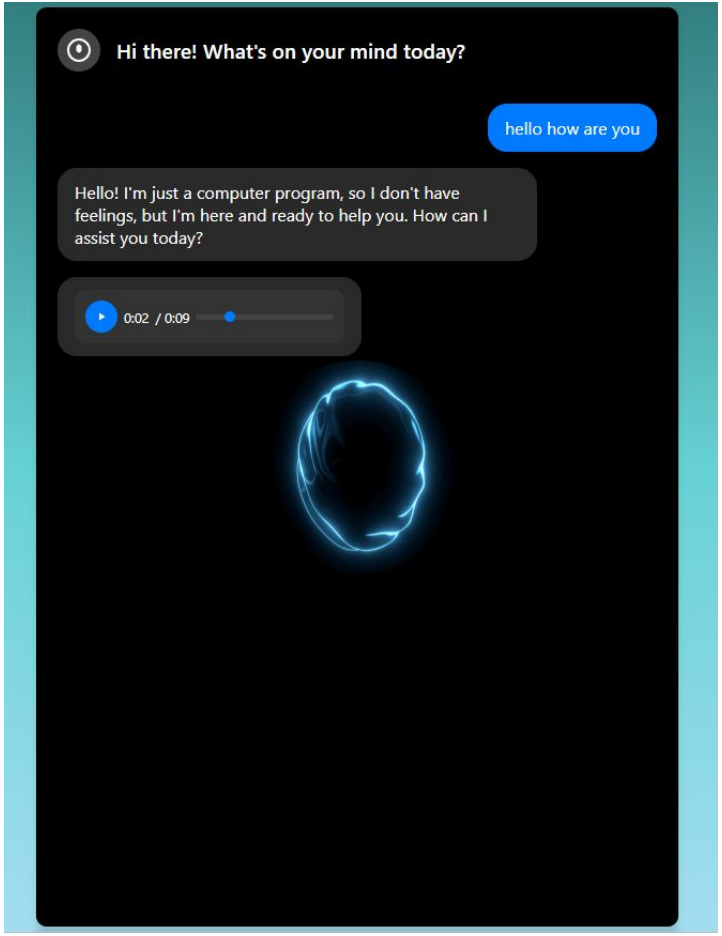
- **Real-Time Interaction**: Immediate transcription and response
- **Seamless Audio Integration**: Easy upload and playback
- **Dynamic TTS Feedback**: Unique response generation for every interaction
- **Scalable Design**: Modular architecture for future improvements

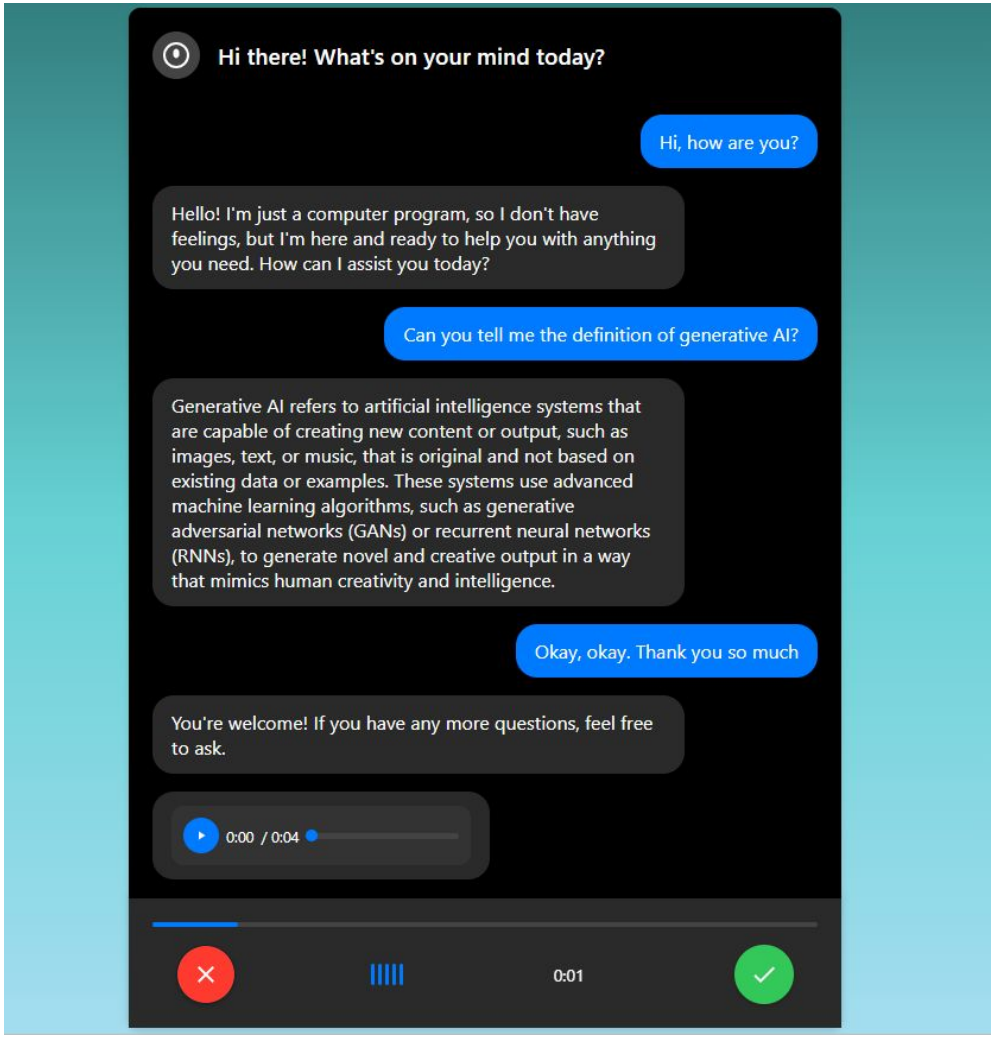
```python
def generate_tts(text, output_dir="static/audio"):
    global previous_audio_file
    os.makedirs(output_dir, exist_ok=True)

    # Generate a unique filename for the audio file
    unique_filename = f"response_{uuid.uuid4().hex}.mp3"
    # unique_filename = f"response.mp3"
    filepath = os.path.join(output_dir, unique_filename)

    # Generate and save the TTS audio
    tts = gTTS(text=text, lang="en")
    tts.save(filepath)
    # Delete the previous file if it exists
    if previous_audio_file and os.path.exists(previous_audio_file):
        try:
            os.remove(previous_audio_file)
        except Exception as e:
            print(f"Error deleting file {previous_audio_file}: {e}")
```

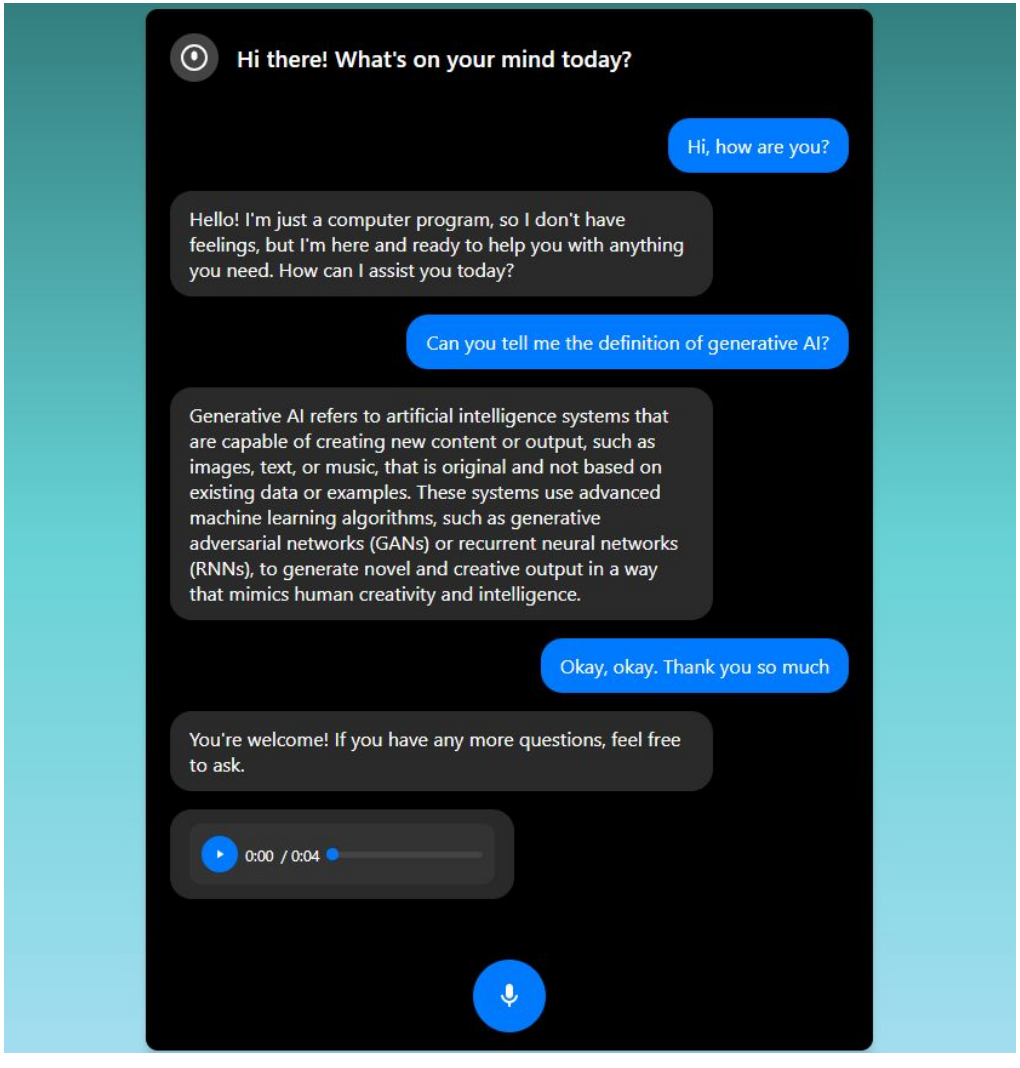
```python
def transcribe_audio(audio_file):
    try:
        # Resolve the absolute path to the `temp` folder
        base_dir = os.path.dirname(os.path.abspath(__file__))  # Get directory of `utility.py`
        temp_dir = os.path.join(base_dir, "temp")  # Resolve the temp directory path

        # Ensure the temp directory exists
        os.makedirs(temp_dir, exist_ok=True)

        # Save the audio file to the `temp` directory
        filename = secure_filename(audio_file.filename)
        temp_path = os.path.join(temp_dir, filename)
        audio_file.save(temp_path)

        audio_file= open(temp_path, "rb")
        transcription = client.audio.transcriptions.create(
        model="whisper-1",
        file=audio_file
        )
        print(transcription.text)
        return transcription.text

    except Exception as e:
        return f"Error during transcription: {str(e)}"
```

Amazon Q Tip 2/3: Invoke suggestions


Hi there! What's on your mind today?


Hi there! What's on your mind today?
hello how are you
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?
0:02 / 0:09


Hi there! What's on your mind today?
Hi, how are you?
Hello! I'm just a computer program, so I don't have feelings, but I'm here and ready to help you with anything you need. How can I assist you today?
Can you tell me the definition of generative AI?
Generative AI refers to artificial intelligence systems that are capable of creating new content or output, such as images, text, or music, that is original and not based on existing data or examples. These systems use advanced machine learning algorithms, such as generative adversarial networks (GANs) or recurrent neural networks (RNNs), to generate novel and creative output in a way that mimics human creativity and intelligence.
Okay, okay. Thank you so much
You're welcome! If you have any more questions, feel free to ask.
0:00 / 0:04
0:01


Hi there! What's on your mind today?
Hi, how are you?
Hello! I'm just a computer program, so I don't have feelings, but I'm here and ready to help you with anything you need. How can I assist you today?
Can you tell me the definition of generative AI?
Generative AI refers to artificial intelligence systems that are capable of creating new content or output, such as images, text, or music, that is original and not based on existing data or examples. These systems use advanced machine learning algorithms, such as generative adversarial networks (GANs) or recurrent neural networks (RNNs), to generate novel and creative output in a way that mimics human creativity and intelligence.
Okay, okay. Thank you so much
You're welcome! If you have any more questions, feel free to ask.
0:00 / 0:04