

GPT Speech Conversation

In this work, we use GPT-4 for general question answering. We feed audio prompts to Whisper-1 model of OpenAI, which converts speech to text. The text is then fed as a prompt to GPT-4 model which generates responses. The text responses are converted to audio responses using library pyttsx3.

Features

GPT-4 is a large multimodal model which accepts text and image inputs and outputs text. Our model is fed text questions, converted from speech using OpenAI's API. Model uses following settings.

```
model="gpt-4",  
messages=messages,  
temperature=0.5,  
max_tokens=100,  
top_p=1.0,  
frequency_penalty=0.0,  
presence_penalty=0.0
```

Capabilities

GPT-4 passed a simulated bar exam with a score around the top 10% of test takers. It can generate responses with 8192 tokens. Our model takes audio prompt and generates audio responses.

Functionality

Functionality has been described in notebook `gpt_speech_conversation.ipynb`