

# Integrating ChatGPT and D-ID AI for a Therapeutic AI Application

## Abstract

This report presents the development and implementation of a therapeutic AI application that combines the power of OpenAI's ChatGPT and D-ID AI. The application aims to assist users in addressing their problems by providing personalized and empathetic responses through a virtual assistant. The user inputs their concerns, and the application responds with a message generated by ChatGPT. Subsequently, this message is transformed into a video response using D-ID AI, creating a more engaging and human-like interaction.

## Introduction

In recent years, AI technologies have shown promise in various applications, including mental health support. Therapeutic applications have gained popularity in providing mental health support. This project explores the integration of two advanced AI technologies, ChatGPT and D-ID, the combination of natural language processing (NLP) and video generation technologies allows us to create a therapeutic AI that responds to users' concerns in a more emotionally resonant manner.

## Research about AI APIs that exist

During my investigation into AI APIs, I encountered the [Blockade Labs API](#), which intrigued me due to its capacity to generate a skybox based on user input. However, my attempts to leverage this API revealed a limitation in the flexibility of parameters available for customization. The constraints in parameter options hindered my ability to fine-tune and tailor the generated skybox to meet the specific needs of the application.

Undeterred, I extended my search to explore AI video generating APIs as an alternative avenue. While I identified several options, a recurring challenge was the prevalence of paid usage models, and the majority offered only limited access with free trials allowing videos of less than one minute in duration.

Despite these initial hurdles, the pursuit of a suitable solution eventually led to the discovery of the D-ID AI, which, with its unique features and capabilities, emerged as the optimal choice for transforming textual responses into engaging and personalized video content.

## Methodology

**User Input Handling:** Users input their problems into the application, which is then sent to the ChatGPT API for generating textual responses.

**Textual Response Generation:** ChatGPT processes the user input and generates empathetic and supportive textual responses, mimicking a therapeutic conversation.

**Video Generation with D-ID AI:** The generated text responses are then sent to the D-ID API, which transforms the text into a video. D-ID specializes in creating lifelike videos that enhance user engagement.

**Displaying Video Responses:** The created video is displayed within the application, offering users a unique and personalized therapeutic experience.

When the user sets the API key for D-ID API a request is made to the D-ID API to create a silent video of the actor which is displayed when the user is telling about how they are feeling. After this the text is sent to open AI and the response is sent to D-ID to create a video of the actor talking, which is displayed on the application. After the video ends, the silent video starts playing again, waiting for the user to respond.

## Special Features

**Emotional Intelligence:** ChatGPT's advanced NLP capabilities enable the chatbot to respond with a high degree of emotional intelligence, providing users with a more empathetic interaction.

**Personalized Video Responses:** The integration with D-ID AI allows for the creation of personalized video messages, adding a visual and emotional layer to the therapeutic interaction.

## Choice of APIs

**ChatGPT by OpenAI:** ChatGPT, based on the GPT-3 architecture, is chosen for its advanced natural language understanding and generation capabilities. It excels in conversational contexts, making it suitable for creating a responsive and empathetic chatbot.

**D-ID AI:** D-ID AI is selected for its video generation capabilities, specifically tailored for face and voice synthesis. This API enhances the user experience by converting text-based responses into personalized video messages.

## Advanced features of D-ID API

D-ID offers a rich array of features and parameters, providing a versatile toolkit for customization. Users can delve into a multitude of options, allowing for a tailored and immersive experience:

**Actor Customization:** The flexibility to change actors allows for diverse and dynamic interactions. Users can select from a range of personas to suit the tone and style of the therapeutic conversation.

**Facial Expressions:** D-ID's capabilities extend to manipulating facial expressions, contributing to the lifelike quality of the generated videos. This feature adds an additional layer of nuance and authenticity to the visual responses.

**Voice and Accent Customization:** Users have the ability to modify not only the voice but also the accent of the actor in the generated videos. This customization further enhances the personalization, catering to individual preferences and comfort.

**Language Selection:** The option to choose the language in which the actor communicates allows for inclusivity and accessibility.

**Audio Input Compatibility:** D-ID goes beyond text input by accommodating audio files as input for video generation. This feature expands the possibilities for communication, allowing users to generate videos in a more varied and dynamic manner.

**Background Color Adjustment:** The ability to change the background color adds a visual dimension to the overall user experience. This feature allows for a customizable and aesthetically pleasing backdrop, contributing to the immersive nature.

The API offers a diverse selection of actors, each bringing a unique personality and style to the conversation. This wealth of customization options empowers users to tailor their AI-driven therapeutic experience according to their preferences and needs, making the application more adaptable and engaging.

Here are some of the actors provided by the API.



# Experimental Results

I conducted a series of experiments using OpenAI's ChatGPT to assess the quality of its responses. In nearly all instances, the responses were favorable. Here are some of the prompts I provided and the corresponding responses generated by ChatGPT:

**User:** I'm feeling sad. Help me uplift my mood.

**ChatGPT:** I'm sorry to hear that you're feeling sad. Here are some tips that might help you uplift your mood: Listening to music can help you feel better by reducing stress and anxiety. You can try listening to upbeat music or your favorite songs to help you feel more positive. Get moving: Exercise is a great way to boost your mood. Even a short walk or light exercise can help you feel more energized and positive.

**User:** I'm feeling sad. Help me uplift my mood.

**ChatGPT:** I'm sorry to hear that you're feeling lonely. Here are some tips that might help you make new friends: Joining a club or group that interests you can be a great way to meet new people who share your interests. You can try joining a book club, a sports team, or a volunteer organization. Taking a class can be a great way to meet new people and learn new skills. You can try taking a cooking class, a language class, or a dance class.

## Problems Encountered

The constraint of limiting ChatGPT's response to two sentences was necessitated by the processing time and credit consumption associated with D-ID, particularly for lengthier text inputs. Additionally, the decision to forego autoplay for video responses was driven by the need to mitigate delays, as autoplay typically awaits the full video loading, resulting in a smoother and more immediate user experience. However, this approach aims to strike a balance between content delivery efficiency and resource optimization within the therapeutic AI application.

## Conclusion

The integration of ChatGPT and D-ID AI in a therapeutic AI application represents a pioneering approach to mental health support. The combination of text and video responses provides users with a more comprehensive and emotionally resonant experience. The positive experimental results underscore the potential of this approach in creating advanced and effective therapeutic AI applications.

This project showcases the potential of combining advanced NLP and video generation technologies to enhance virtual therapeutic experiences.

## Future Work

Future iterations of the application could involve refining the AI models for more advanced responses and exploring additional features such as real-time interaction, multi-language support, and expanded emotional intelligence.

This project showcases the possibilities of combining diverse AI technologies to create innovative solutions for mental health support, paving the way for more advanced and compassionate AI-driven applications in the future.

## References

1. Getting Started 🤖 (d-id.com)
2. Overview - OpenAI API

## Links to the Worlds

1. Therapeutic AI

[World: Therapeutic AI by Moses Crasto \[Ancient Brain\]](#)

2. D-ID API

[World: D-ID\(clone by Moses Crasto\)\[Ancient Brain\]](#)

3. openAI API

[World: Chat with GPT model \(clone by Moses Crasto\)\[Ancient Brain\]](#)