**Familiar**

**A personalized sentient Avatar for self-help**

**Problem Statement**

At the heart of this endeavour lies the profound challenge of addressing the intricate emotional aftermath following the loss of a loved one—an experience often intertwined with the onset of conditions like Post-Traumatic Stress Disorder (PTSD) and depression. Although existing therapeutic modalities endeavour to offer relief, they grapple with the formidable task of encompassing the multifaceted range of emotions and intricacies entangled in the fabric of grief.

**Problem Solution**

Our solution to the said problem is an application that comprises a personalized sentient Avatar that is used for dealing with PTSD. An avatar is an artificial intelligence (AI) chatbot that can not only converse with users using natural language but also generate and understand images based on the dialogue context. A visual chatbot that talks like a person can provide a more engaging and realistic experience for users and enhance the communication and expression of ideas through visual media.

**Objectives**

The main objectives of this project are:

* **Avatar Development:** Design and develop a personalized sentient Avatar that can be customized with the image of a departed loved one, effectively providing a virtual presence and connection for users.
* **AI-driven Sentiment Analysis:** Implement advanced AI algorithms to analyze user interactions and sentiments, enabling the Avatar to understand emotional cues and tailor its responses accordingly.
* **Personalized Motivational Content:** Curate a repository of motivational content, including quotes, anecdotes, and positive affirmations, that the avatar can share with users based on their emotional cues.
* **Responsive User Engagement:** Develop an interface that encourages users to engage regularly with the avatar, fostering a consistent source of emotional support and motivation.

**Challenges**

The main challenges of this project are:

* **Emotion Detection and Analysis:** Developing accurate algorithms to detect and analyze emotional content in user inputs, which can be complex given the diversity and subtlety of human emotions.
* **Real-time Interaction:** Enabling real-time interactions with minimal latency while maintaining a seamless and responsive conversation flow between users and the avatar.
* **Avatar Customization:** Implementing a user-friendly system to allow users to customize the avatar's appearance and attributes, and ensuring that the customization process is intuitive and visually appealing.
* **Emulating Empathy:** Developing an avatar with genuine empathetic responses that are meaningful to users poses a technological challenge. The avatar needs to communicate empathy without giving the impression of insincerity.
* **Stigma and Acceptance:** Addressing the potential stigma around seeking support from a virtual avatar rather than traditional sources can impact user adoption. Promoting acceptance and understanding of the avatar's role is a challenge.

**Scope**

This project encompasses the development of a comprehensive conversational AI system divided into four distinct modules. Each module is designed to address specific aspects of the project, aiming to provide users with a seamless and personalized experience.

The project's scope is defined through the following subsections:

**Modules:**

**Module 1: Chatbot Development and Integration**

This module focuses on developing a functional chatbot system that employs the ChatGPT API from the langchain library. The GPT-3.5 Turbo model will be integrated to facilitate natural language understanding and generation. The chatbot's context preservation mechanism and user conversation history storage will ensure personalized interactions. Users can choose context lengths, with pricing based on token usage. Whisper TTS model integration will enable voice-to-text conversion, while a web application will serve as the user interface.

**Models:**

This module will employ the *ChatGPT API* from the *langchain* library, integrating the *GPT-3.5 Turbo* model. This model will serve as the foundation for natural language understanding and generation in the chatbot system.

**Pricing:**

There are two price points available to use the *GPT-3.5 Turbo* model based on the amount of history and context you can store:

* 4K context: $0.0015 per 1K tokens input, $0.002 per 1K tokens output
* 16K context: $0.003 per 1K tokens input, $0.004 per 1K tokens output

**Datasets:**

The primary dataset requirement lies within the *ChatGPT API* and the *Whisper TTS* model for voice-to-text conversion. No additional external datasets are needed for this module.

**Module 2: Avatar Creation and Expression Control**

This module explores methods to generate avatars from user-provided images. The objective is to create lifelike avatars that can exhibit a range of expressions. The approach may involve researching existing algorithms or adapting *GAN-based* techniques. This module also contemplates the need for datasets for potential training purposes.

**Models:**

This module will involve the exploration of existing algorithms or models for generating avatars from user-provided images. Research will be conducted to identify and possibly adapt *GAN-based* approaches or similar methods for creating lifelike avatars.

**Pricing:**

Costs for this module are predominantly related to research and development efforts, as no predefined external models are intended for integration.

**Datasets:**

Depending on the chosen approach, datasets may be needed for training purposes. This could involve facial image datasets for *GAN-based* techniques or other relevant data sources.

**Module 3: Personalized Speech Synthesis**

This module aims to identify models capable of converting text-based chatbot responses into personalized speech. The project will consider existing open-source models such as *Tacotron* and *DeepVoice*, evaluating their ability to match user-provided voices and generate natural speech. Potential datasets containing speech data and text transcripts will be explored for training and customization purposes.

**Models:**

This module aims to identify suitable models for converting text-based chatbot responses into personalized speech. Existing open-source models like *Tacotron* and *DeepVoice* will be considered for their potential to match user-provided voices and produce natural speech.

**Pricing:**

Costs for this module encompass research, experimentation, and potential model integration. The choice of an open-source solution will influence the associated costs.

**Datasets:**

Should model training be pursued, datasets containing speech data and corresponding text transcripts will be required to facilitate training and customization.

**Module 4: Avatar-Response Integration and Refinement**

In this module, various models will be assessed for syncing chatbot responses with avatar actions, encompassing facial expressions, lip-syncing, and emotional cues. Models like speech2face and lip-syncing techniques will be tested, along with emotion recognition models like DeepFace and EmoPy. The goal is to seamlessly blend text and speech responses with avatar movements for a lifelike interaction.

**Models:**

In this module, a variety of models will be evaluated to sync chatbot responses with avatar actions. Models include *speech2face*, *lip-syncing techniques,* and *emotion recognition models* such as *DeepFace* and *EmoPy*.

**Pricing:**

This module's costs involve testing, fine-tuning, and model integration. Licensing fees or access costs for selected external models will also be considered.

**Datasets:**

Potential datasets might be necessary for training models related to lip-syncing, emotion recognition, and facial expression generation to enhance synchronization accuracy.

**Exclusions**

The project excludes:

* Advanced AI models beyond the chosen solutions
* Complex user authentication features
* In-depth model training beyond the project's scope
* Response Customization of the Avatar

**Boundaries**

Project limitations include:

* Adherence to the capabilities of selected AI models and APIs
* Resource availability for research, development, and testing
* Budget constraints for model research and integration

**Assumptions**

Assumptions made for the project include the successful integration of selected AI models, the availability of open-source solutions for personalized speech synthesis, and the adaptation of avatar generation techniques.

**Constraints**

The project operates under the constraints of budget, time, and available resources for research, development, and integration efforts.

**Dependencies**

The project's success depends on the availability and integration of external AI models such as *ChatGPT*, *Whisper TTS*, and selected models for avatar synchronization and personalized speech synthesis.

**Risks**

Potential risks include API downtime affecting real-time interactions, challenges in avatar creation, and limited availability of open-source speech synthesis models.

**Feasibility**

This feasibility analysis aims to assess the practicality and viability of the proposed project in terms of technical, economic, and operational factors. Each module's feasibility will be evaluated to ensure a successful implementation:

**Technical Feasibility**

**Module 1: Chatbot Development and Integration**

The technical feasibility of this module is high, as the ChatGPT API and GPT-3.5 Turbo model are readily available and well-documented. The integration process has been accomplished in various projects. The Whisper TTS model's integration is also technically feasible, given the availability of APIs and tools for text-to-speech conversion. The primary challenge lies in implementing the conversation history mechanism for personalized interactions.

**Module 2: Avatar Creation and Expression Control**

The technical feasibility of this module is moderate. While existing algorithms and GAN-based approaches for avatar creation are available, the adaptation and integration of these methods may require technical expertise. Dataset availability and image processing capabilities are key factors that will influence technical feasibility.

**Module 3: Personalized Speech Synthesis**

The technical feasibility of this module is moderate. Integrating existing speech synthesis models like Tacotron or DeepVoice is technically feasible, but the challenge lies in matching user-provided voices and generating natural speech. Technical limitations might arise in the customization of speech models to ensure a personalized experience.

**Module 4: Avatar-Response Integration and Refinement**

The technical feasibility of this module is challenging. Synchronizing chatbot responses with avatar actions, including lip-syncing and facial expressions, demands complex model integration. While individual models exist for various aspects, achieving seamless synchronization will require substantial technical effort.

**Economic Feasibility**

**Module 1: Chatbot Development and Integration**

The economic feasibility of this module is high, as the integration of ChatGPT and Whisper TTS models is cost-effective due to their accessible APIs. The main cost considerations include API usage fees and potential development costs for the web application.

**Module 2: Avatar Creation and Expression Control**

The economic feasibility of this module is moderate. The cost involves research, development, and potential dataset acquisition. If no external models are integrated, costs remain manageable.

**Module 3: Personalized Speech Synthesis**

The economic feasibility of this module is moderate to high. If open-source models are available, costs will be limited to research, experimentation, and potential integration efforts. If model training is required, costs may increase due to dataset procurement and training efforts.

**Module 4: Avatar-Response Integration and Refinement**

The economic feasibility of this module is moderate to high. Testing and integrating various models will involve expenses. Licensing fees or access costs for external models may influence costs.

**Operational Feasibility**

**Module 1: Chatbot Development and Integration**

The operational feasibility of this module is high. The ChatGPT API and Whisper TTS models are user-friendly and well-documented, enabling efficient integration. The development of a web application for user interaction is a common operational task.

**Module 2: Avatar Creation and Expression Control**

The operational feasibility of this module is moderate. Operational challenges may arise during the research and development phase, requiring specialized skills in image processing and avatar generation.

**Module 3: Personalized Speech Synthesis**

The operational feasibility of this module is moderate. If suitable open-source models are available, integration can be streamlined. However, operational challenges may arise if model training is required.

**Module 4: Avatar-Response Integration and Refinement**

The operational feasibility of this module is moderate. Testing and refining various models will require technical expertise. Operational challenges may arise when integrating multiple models for synchronization.

**References**

<https://github.com/saiteja-talluri/Speech2Face.git>

<https://github.com/aqibahmad/speech2face.git>

<https://looking-to-listen.github.io/avspeech/download.html>

[AVSpeech: Audio Visual Speech dataset (looking-to-listen.github.io)](https://looking-to-listen.github.io/avspeech/download.html)

<https://github.com/googlesamples/mediapipe/tree/main/examples/face_landmarker/python>

[Face landmark detection guide for Python | MediaPipe | Google for Developers](https://developers.google.com/mediapipe/solutions/vision/face_landmarker/python)

[GitHub - Rudrabha/LipGAN: This repository contains the codes for LipGAN. LipGAN was published as a part of the paper titled "Towards Automatic Face-to-Face Translation".](https://github.com/Rudrabha/LipGAN)

[lip-sync · GitHub Topics · GitHub](https://github.com/topics/lip-sync)

[facial-expression-recognition · GitHub Topics · GitHub](https://github.com/topics/facial-expression-recognition?l=python)