



FAST NUCES, ISLAMABAD

Familiar

A personalized sentient Avatar for self-help



Syed Addan 20I-0818
Hasam Nazir 20I-2438
Faizan Mehmood 20I-2704

supervised by
Mr. SHOAIB SALEEM KHATTAK

August 28, 2023



Contents

1	Problem Statement	2
2	Problem Solution	2
3	Scope	2
3.1	Modules	2
3.1.1	Module 1: Chatbot Development and Integration	2
3.1.2	Module 2: Avatar Creation	3
3.1.3	Module 3: Personalized Speech Synthesis	3
3.1.4	Module 4: Avatar-Response Integration and Expression Control	3
3.2	Exclusions	3
3.3	Boundaries/Limitations	4
3.4	Assumptions	4
3.5	Constraints	4
3.6	Dependencies	4
3.7	Risks	4
4	Feasibility	4
4.1	Technical Feasibility	4
4.1.1	Module 1: Chatbot Development and Integration	4
4.1.2	Module 2: Avatar Creation	4
4.1.3	Module 3: Personalized Speech Synthesis	5
4.1.4	Module 4: Avatar-Response Integration and Expression Control	5
4.2	Economic Feasibility	5
4.2.1	Module 1: Chatbot Development and Integration	5
4.2.2	Module 2: Avatar Creation	5
4.2.3	Module 3: Personalized Speech Synthesis	5
4.2.4	Module 4: Avatar-Response Integration and Expression Control	5
5	Requirements	5
5.1	Functional Requirements	5
5.1.1	Module 1: Chatbot Development and Integration	5
5.1.2	Module 2: Avatar Creation	6
5.1.3	Module 3: Personalized Speech Synthesis	6
5.1.4	Module 4: Avatar-Response Integration and Expression Control	6
5.2	Non-Functional Requirements	7
5.2.1	Module 1: Chatbot Development and Integration	7
5.2.2	Module 2: Avatar Creation	7
5.2.3	Module 3: Personalized Speech Synthesis	7
5.2.4	Module 4: Avatar-Response Integration and Expression Control	8
6	Bonus Work	8
6.1	Enhanced Avatar Customization	8
6.2	Multi-Language Support	8
6.3	Gamified Progress Tracking	8
7	Further Work	8
7.1	Advanced Emotional Analysis	8
7.2	Context-Aware Interactions	8
7.3	User Community and Peer Support	8
7.4	Integration with Professional Therapists	8
7.5	Long-Term Emotional Tracking	9
7.6	Voice Modulation and Personalization	9
7.7	Cross-Platform Accessibility	9
7.8	Integration with Wearable Devices	9

1 Problem Statement

Grieving the loss of a loved one is a profound and very personal process that typically brings forth a spectrum of complex emotions. This emotional aftermath can result in disorders such as Post-Traumatic Stress Disorder (PTSD) and depression, posing considerable obstacles to persons' well-being. While professional therapy can be beneficial, people dealing with these problems frequently withdraw into solitude, missing out on important human interaction and expert counsel. Existing therapeutic methods, while well-intended, may struggle to handle the complexities of grieving and the necessity of getting help adequately.

2 Problem Solution

Introducing **Familiar**, a groundbreaking application designed to illuminate the paths of healing and connection for individuals navigating the storm of emotions following the loss of a loved one. By harnessing cutting-edge technology, *Familiar introduces a personalized sentient companion*, uniquely crafted to provide a sense of virtual companionship during times of emotional turmoil. This empathetic companion, seamlessly integrated with *AI-powered sentiment analysis* and *responsive interactions*, becomes a trusted confidant, capable of offering tailored emotional support.

Familiar gently empowers individuals to step out from the shadows of isolation and embrace the healing power of human connection. By engaging in candid conversations, the companion encourages users to consider seeking professional assistance, fostering an understanding that there's strength in sharing one's struggles. Beyond mere virtual interactions, the application aims to spark a renewed sense of hope, motivating users to explore ways of healing and support that transcend the digital realm.

In an era where technology can both isolate and unite, Familiar seeks to strike a harmonious balance. It acknowledges the crucial role of genuine human connection and professional guidance in the healing process. Through Familiar, we aim to rekindle the flame of resilience within each user, guiding them towards a path of emotional recovery and enriching their lives with the understanding that seeking help is a courageous and transformative choice. **Embrace Familiar, your personal sentient Avatar for self-help, and embark on a journey of healing and renewal.**

3 Scope

This project encompasses the development of a comprehensive conversational AI system divided into four distinct modules. Each module is designed to highlight high-level features, aiming to provide users with a seamless and personalized experience.

3.1 Modules

3.1.1 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 personalised interactions. Users can choose context lengths, with pricing based on token usage. *Whisper STT 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. The chatbot will have guardrails in place to prevent the users to inputting inappropriate inputs and will be instructed beforehand to give responses that are motivational and self-help related. And also the *Whisper Speech to Text model* will be used to convert the users voice input to text and then the text will be paced into the chatbot created before to get a response. The users will have a *Web Application* as a User Interface to interact with.

Datasets No additional external datasets are needed for this module.

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 dollars per 1K tokens input, 0.002 dollars per 1K tokens output
- 16K context: 0.003 dollars per 1K tokens input, 0.004 dollars per 1K tokens output

3.1.2 Module 2: Avatar Creation

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

Datasets

Pricing

3.1.3 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*, *Real-Time-Voice-Cloning* 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

Datasets

Pricing

3.1.4 Module 4: Avatar-Response Integration and Expression Control

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

Datasets

Pricing

3.2 Exclusions

Familiar won't include the following:

- Advanced AI models beyond the chosen solutions
- In-depth model training beyond the project's scope
- Response Customization of the Chatbot
- Avatar Customization
- Any response from the Chatbot beyond the guardrails we have defined
- Complex user authentication features

3.3 Boundaries/Limitations

Familiar will adhere to the following limitations:

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

3.4 Assumptions

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

3.5 Constraints

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

3.6 Dependencies

a'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*.

3.7 Risks

Potential risks include in the completion of **Familiar**:

- API downtime affecting real-time interactions
- Challenges in avatar creation
- Scarceness of required resources
- Limited availability of open-source models

4 Feasibility

This feasibility analysis aims to assess the practicality and viability of **Familiar** in terms of **technical and economic factors**. Each module's feasibility will be evaluated on a **scale of easy-moderate-challenging** based on the resources available for it, the technical aspects of it and the time required to implement it.

4.1 Technical Feasibility

4.1.1 Module 1: Chatbot Development and Integration

The technical feasibility of this module is easy, 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 speech-to-text conversion. The primary challenge lies in implementing the conversation history mechanism for personalized interactions as that will cost us too. The **Web Application development** is also very feasible as they are numerous resources to help us through the process of its creation.

4.1.2 Module 2: Avatar Creation

The technical feasibility of this module is challenging. *Existing algorithms* and *GAN-based approaches* for avatar creation are available but none match our requirements, so we will have to **train our own model**. **Various Datasets** are available to train the model and **hardware is assumed to be available**.

4.1.3 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.

4.1.4 Module 4: Avatar-Response Integration and Expression Control

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

4.2 Economic Feasibility

4.2.1 Module 1: Chatbot Development and Integration

The economic feasibility of this module is moderate, because though the integration of ChatGPT and Whisper TTS models may be cost-effective due to their accessible APIs but we will have to buy the tokens for *ChatGPT 3.5 Turbo*'s API as mentioned in the Module's scope. The main cost consideration is API usage fees.

4.2.2 Module 2: Avatar Creation

The economic feasibility of this module is moderate. The cost involves fulfilling hardware requirements to train the model as there were no models that could have our project's requirements.

4.2.3 Module 3: Personalized Speech Synthesis

The economic feasibility of this module is easy, as open-source models such as *Real-Time-Voice-Cloning* are available.

4.2.4 Module 4: Avatar-Response Integration and Expression Control

The economic feasibility of this module is moderate. The cost involves fulfilling hardware requirements to train the model as there were no models that could have our project's requirements.

5 Requirements

This section specifies the different needs and requirements of the users and the stakeholders that they have with **Familiar**. This section outlines what the project aims to achieve in terms of **scope, deliverables, and performance**. It provides a clear and detailed description of what needs to be developed, designed, or implemented to successfully complete the project. There are two types of requirements, **Functional and Non-Functional Requirements**.

5.1 Functional Requirements

5.1.1 Module 1: Chatbot Development and Integration

1. Chatbot System Development

- Develop a functional chatbot system using the ChatGPT API from the langchain library
- Integrate the GPT-3.5 Turbo model for natural language understanding and generation
- Implement context preservation to store user conversation history

2. User Interaction

- Enable users to interact with the chatbot through natural language inputs
- Provide options for users to select context lengths for conversations

3. Voice-to-Text Conversion

- Integrate the Whisper Speech to Text (STT) model for voice-to-text conversion

4. Web Application Interface

- Develop a web application to serve as the user interface for interacting with the chatbot

5. Guardrails and Responses

- Implement mechanisms to prevent inappropriate inputs from users
- Configure the chatbot to respond with motivational and self-help content

6. Pricing

- Offer two pricing tiers based on the amount of history and context users can store
- Calculate pricing for token input and output based on the selected context size

5.1.2 Module 2: Avatar Creation

1. Avatar Generation

- Research and explore methods to generate lifelike avatars from user-provided images
- Investigate GAN-based techniques and existing algorithms for avatar creation

2. Expression Range

- Develop avatars capable of exhibiting a wide range of expressions

3. Datasets

- Explore potential datasets for training avatar generation models

5.1.3 Module 3: Personalized Speech Synthesis

1. Text-to-Speech Conversion

- Identify and evaluate models such as Tacotron, Real-Time-Voice-Cloning, and DeepVoice for converting text-based chatbot responses into personalized speech

2. Voice Matching

- Assess the capability of selected models to match user-provided voices and generate natural speech

3. Training Data

- Research and explore datasets containing speech data and corresponding text transcripts for training and customization

5.1.4 Module 4: Avatar-Response Integration and Expression Control

1. Avatar-Response Synchronization

- Evaluate models like speech2face and lip-syncing techniques for synchronizing chatbot responses with avatar actions

2. Emotion Recognition

- Test emotion recognition models such as DeepFace and EmoPy for capturing emotional cues in avatars

3. Realistic Interaction

- Blend text and speech responses seamlessly with avatar movements to create lifelike interactions

5.2 Non-Functional Requirements

5.2.1 Module 1: Chatbot Development and Integration

1. Usability

- The user interface of the chatbot system should be intuitive and easy to use, ensuring a positive user experience
- The web application interface should be responsive and user-friendly across different devices and screen sizes
- Users should be able to interact with the chatbot and access features without requiring extensive training or technical knowledge

2. Performance

- The chatbot responses should be generated and displayed within a maximum of 3 seconds to maintain a natural conversation pace
- The voice-to-text conversion using the Whisper STT model should have low latency to enable smooth voice interactions
- The web application should load and respond quickly to user interactions, ensuring minimal waiting times

3. Security and Privacy

- User interactions and data should be encrypted during transmission to ensure data security
- The user conversation history and personal information should be stored securely and protected from unauthorized access
- The chatbot system should adhere to data privacy regulations and best practices to safeguard user information

5.2.2 Module 2: Avatar Creation

1. Realism and Quality

- The generated avatars should exhibit a high degree of realism in facial expressions and movements, providing an immersive experience
- Avatars should accurately reflect the emotional cues conveyed through chatbot responses to ensure synchronization

2. Efficiency

- The avatar generation process should be efficient and not cause significant delays in user interactions
- Computational resources should be optimized to prevent excessive resource consumption during avatar creation

5.2.3 Module 3: Personalized Speech Synthesis

1. Speech Naturalness

- The synthesized speech should sound natural and human-like, ensuring a high-quality listening experience for users
- Voice synthesis models should be capable of conveying emotional nuances present in the text-based responses

2. Speed

- Text-based chatbot responses should be converted to personalized speech within 2 seconds to maintain a fluid conversation flow
- The speech synthesis process should not cause significant delays in the overall interaction

3. Resource Utilization

- Optimize computational resources during speech synthesis to ensure efficient processing without overburdening the system

5.2.4 Module 4: Avatar-Response Integration and Expression Control

1. Real-time Interaction

- The synchronization between chatbot responses and avatar actions should occur in real-time to create a lifelike and immersive experience
- The delay between chatbot responses and corresponding avatar expressions should not exceed 1 second

2. Accuracy of Expression

- Avatar expressions should accurately reflect the emotional and facial cues from chatbot responses, maintaining consistency
- The lip-syncing and emotion recognition techniques should achieve a minimum accuracy rate of 80 percent

6 Bonus Work

6.1 Enhanced Avatar Customization

Explore the possibility of allowing users to customize the appearance and expressions of their avatars. This could involve developing an interface that enables users to modify features such as facial expressions, hair, clothing, and accessories

6.2 Multi-Language Support

Extend the chatbot's capabilities to support multiple languages. Implement language detection and translation mechanisms to provide a personalized experience for users regardless of their language preference

6.3 Gamified Progress Tracking

Introduce gamification elements to track users' progress in their emotional well-being journey. Implement reward systems, achievement badges, and progress tracking charts to motivate users to engage consistently with the application

7 Further Work

7.1 Advanced Emotional Analysis

Explore advanced sentiment analysis techniques to better understand users' emotional states. Investigate machine learning models that can detect subtle emotional cues from text and voice inputs, allowing the application to respond more empathetically

7.2 Context-Aware Interactions

Develop mechanisms to make the chatbot's interactions more context-aware. Implement memory mechanisms that enable the chatbot to remember past conversations and refer back to them, creating a more personalized and coherent conversation flow

7.3 User Community and Peer Support

Consider incorporating features that facilitate interactions among users who are going through similar experiences. Develop discussion forums or support groups where users can share their stories, provide encouragement, and exchange coping strategies

7.4 Integration with Professional Therapists

Investigate partnerships with licensed therapists to integrate their expertise into the application. Implement mechanisms that allow users to seamlessly transition from the virtual companion to professional therapy sessions if needed

7.5 Long-Term Emotional Tracking

Implement features that track users' emotional progress over an extended period. Develop data visualization tools that help users observe their emotional journey, identify patterns, and recognize improvements

7.6 Voice Modulation and Personalization

Extend personalized speech synthesis to mimic the user's voice or that of a loved one. Implement voice modulation techniques to make the interactions even more familiar and comforting

7.7 Cross-Platform Accessibility

Extend the application's accessibility to multiple platforms, including mobile devices, tablets, and desktops. Develop native mobile apps for iOS and Android, ensuring a seamless experience on various devices

7.8 Integration with Wearable Devices

Investigate the integration of wearable devices, such as smartwatches or fitness trackers, to gather real-time biometric data. Utilize this data to enhance the chatbot's responses and provide more accurate emotional support