Software Requirement Specification Document for Digital-Twin Creation of Human

Karim ahmed, Khaled Amr, Martina Raafat, Marwan Fathy
Supervised by: Prof. Khaled Hussein
Assistant Supervisor: Sarah Hatem

February 8, 2024

Table 1: Document version history

Version	Date	Reason for Change
1.0	4-Jan-2024	SRS First version's specifications are defined.
1.1	7-Jan-2024	Intro,Functional Req.,System Description edited
1.2	10-Jan-2024	Add UML Diagram
1.3	11-Jan-2024	Edited system Description
1.4	12-Jan-2024	Add ERD and Premilinary Object-Orianted Domain
1.5	14-Jan-2024	Review Document

GitHub: https://github.com/kkarimwahba/Digital-Twin-creation-of-human

Figure 1: GitHub QR code

Contents

Abstract

The chat-bot has attracted the sights of the market need in the past three years. The world faced a pandemic that led everything to be virtualized. Our objective is to create a personalized and interactive 3D avatar by integrating various deep learning models. Each model serves a specific purpose in achieving different aspects of the avatar. These aspects include image, voice, and text, which are addressed through the use of speech recognition, a 3D avatar generator, and an LLM (Large Language Model), respectively. The user has the ability to generate the avatar by uploading an image that is processed to generate the avatar's body and face, capturing or uploading an audio recording for cloning the user's voice, and engaging in a conversation with the LLM (Large Language Model) model For generating the responses on the user's input.This will lead for generating .

1 Introduction

1.1 Purpose of this document

The goal of digital twin SRS document is to illustrate detailed documentation of the digital twin project. Digital twin is a virtual avatar mobile-based application on which people can upload an image of themselves or for the people who would love to see and that will show similar characteristics to the person of the image and interact with them. Firstly, the purpose and scope of this document will be provided. Followed by an overall explanation of the Digital twin system. Moreover, system functionality such as uploading images of the user, and taking the characteristics. Finally, the used dataset is Alpaca.

1.2 Scope of this document

The document's scope is to tackle similar system to Digital twin academically and business-wise, moreover illustrate the system overview, scope, and context. Also the objectives of the Digital twin mobile-based application, and user characteristics are discussed. Furthermore, this document goes through the functional and non-functional requirements of the Digital twin system, the data design, and the object oriented class diagram. Finally, this document also covers the operational scenarios of the system.

1.3 Business Context

There is no doubt that in the last three years chat-bot platforms become very essential in every field possible. According to[1] there are several ways that we can use that chat-bot in many ways. For example process Improvement within an organization chat-bots can assist with internal support systems. Also, at Innovation and Integration chat-bots can be utilized for innovation management, contributing to the company's innovation and customer service. We consider that the client always searches for the easiest way to solve problem or to just taking support to do something, what would be better than chatting or doing things with a thing that can help you do anything you want when you are at home avoiding a lot of time loss of searching in a lot odd resources, and unsuitable resources without giving you a perfect example or idea of what you need. Integrating chat-bots

into the business will increase the efficiency of every customer also, as known chat-bots can gather valuable collection insights into customer behavior and preferences which can handle a lot of issues for the customer. These benefits give a value to the two sides the customer and chat-bots that gather the information and store it.

2 Similar Systems

2.1 Academic

In, **QLORA**[2], The main goal of the QLORA method is to preserve excellent performance levels while using less memory. To train and assess their models, they used multiple datasets, such as crowdsourced data, distillation from models with fine-tuned instructions, and corpora aggregations. Also, they used a quantized optimizer to update the weights during training, the QLORA technique quantizes the language model's weights. They used (Paged Optimizers and Double Quantization) to represent the higher accuracy. The accuracy of each model is discussed in the study, with special attention paid to the 33B Guanaco model's performance, which can be trained in less than 12 hours on consumer GPUs with 24 GB of RAM. The output includes proof of the QLORA approach's efficacy in creating cutting-edge chatbots that can compete with established models like ChatGPT. they also highlight the possibility of using QLORA tuning on specific open-source data in future work, which can be competitive with the best commercial models now available on the market.

In, **Attention Is All You Need** [3], they discuss the novel network architecture for sequence transduction tasks based only on attention mechanisms. In machine translation applications, the Transformer model performs better than current models built on complex recurrent or convolutional neural networks. They used the standard WMT 2014 English-German dataset and for English-French, they used the larger WMT 2014 English-French dataset. They employed byte-pair encoding for English-German and word-piece vocabulary for English-French. The accuracy is presented in the research along with a comparison of their translation quality and training costs to other model architectures found in the literature. For inference, they employed beam search with a beam. Additionally, they adjusted their base model in several ways to determine the importance of different Transformer components by analyzing the change in translation performance from English to German. The optimizer (an Adam optimizer with specific parameters), the regularization strategies used, the hardware and schedule used for training, and other details of the training regime are all covered in length . They also talk about the models' interpretability, focusing on how self-attention may create better interpretable models.

In, **A survey of Large Language Model**[4], This provides a comprehensive review of (LLM), focusing on models larger than 10B. It covers findings, key concepts, and techniques for understanding and utilizing LLM, including pre-trained, evaluation, adaptation, and utilization, They also address the challenges and future directions for LLMs and their process. All of this is to ensure user privacy and data security. The review includes experimental results and findings on closed-source models as well as improving the use of LLMs in classic NLP tasks. Overall it serves valuable reference for engineers and researchers interested in LLMs and provides insights about their capabilities, challenges and future direction. Also, the paper discuss the detailed of dataset used to evaluate the performance of LLM(Large Language Model) in various NLP tasks. The paper also, reports the zero-shot performance of LLM on these datasets which means that models are evaluated on tasks they havn not trained on. This evaluation metrice reflect the ability of LLMs to generalize new tasks. Overall, the use of these datasets provide valuable insights into the capabilities of LLMs in various NLP tasks.

2.2 Business Applications

2.2.1 Replika Chat-Bot

Replika is an AI chatbot designed to engage with users through a conversation. Replika is for anyone who wants an artificial friend with no judgment or social anxiety. This app offers users a supportive conversation without judging the user allowing them to express their thoughts and share experiences. [5]

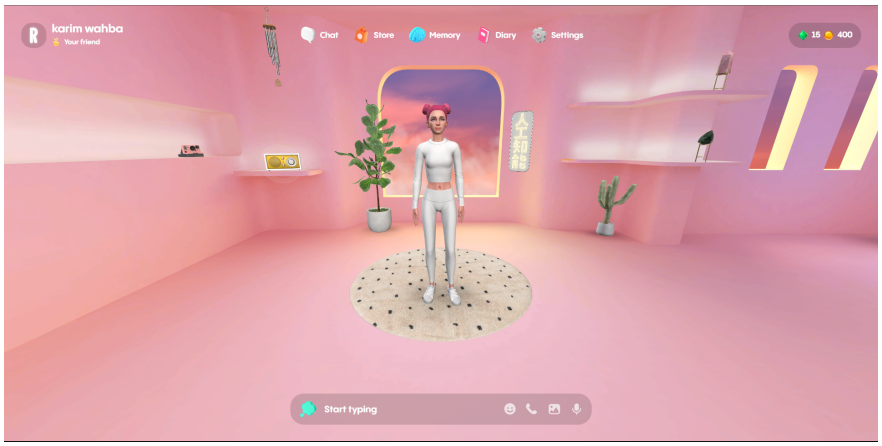

Figure 2: Replika App chat-bot

2.2.2 Read Player Me

Ready Player Me website enables users to create personalized 3D avatars using a simple selfie, which can then be used across a wide range of over 8,000 compatible apps and games. It able the users to customize their avatar such as hairstyle, face, and accessories. [6]

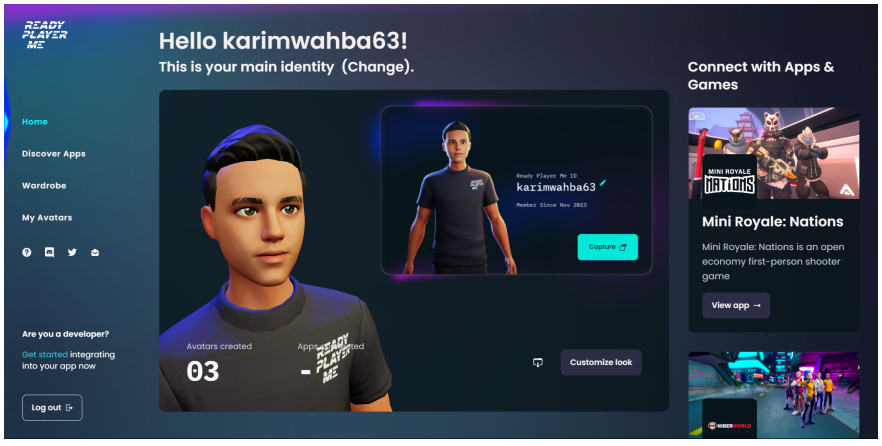

Figure 3: ReadyPlayerMe.com

3 System Description

3.1 Problem Statement

In the existing applications designed to have companionship or simulate the interactions with inaccurate face of the avatar due to limitation for a meaningful and fulfilling experience. The insufficiency of these applications that they cannot deliver the immersive experience in the conversations between the user and the avatar when the user search for the digital twin presence that reminds them of loved ones or a lost wisdom from a mentor. The main problem is the lack of customization of the avatar (digital-twin) options and prevents the user from tailoring the avatar closely to the desired look. Moreover, these applications failed to providing an immersive interaction to conversing with the mentor or the loved one (desired avatar). Our project aims to introduce an innovate system that enhance the user experience that overcome the limitation of the existing applications to enhance the experience of the user when initialing a conversation with the digital twin avatar by offering a high-level of the responsiveness, customization and the interactivity. Providing for the user the ability to chat with their mentor or their loved one whom pass out to re-create memories with them, perform a real-life conversation and helping the user in their daily tasks.

3.2 System Overview

To create an avatar with the same characteristics, viewpoint and voice. we have used a Large Language Model (LLM), which is a type of artificial intelligence model that is trained on large amounts of text data to generate a language similar to human language. It is capable of doing many different tasks related to natural language processing, such as question answering, language translation, and text summarization. LLM consists of the T5, BERT, and GPT-3. Thus, the user must record his voice input and then go to the LLM which includes ASR, used to generate a text from speech, NLP used to allow computers to understand, generate, and collaborate with human language. Then we have Transformer which is a type of neural network architecture consisting of an encoder (to take a text as an input) and a decoder (to generate a text as an output). After that, we applied voice cloning to give the avatar the user's voice.

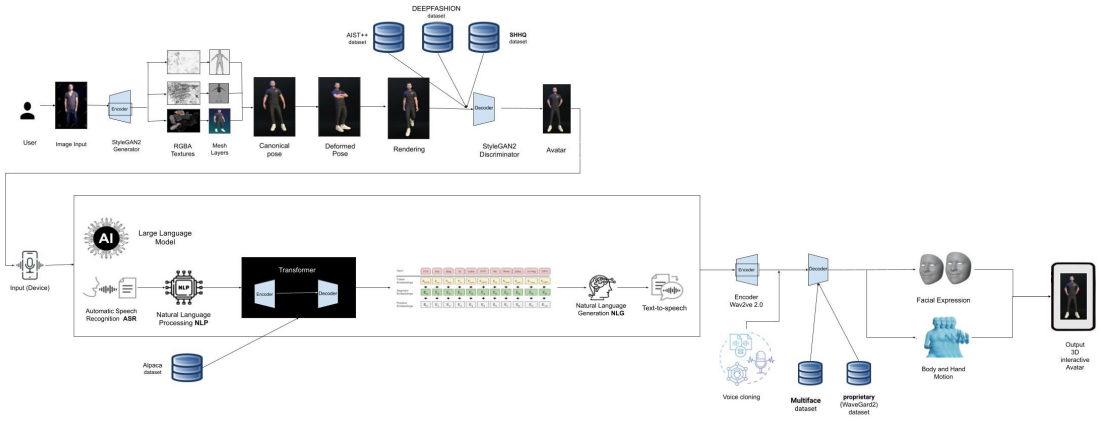

Figure 4: System Overview

3.3 System Scope

Our project aims to create a digital twin that could be the user replacement or a replacement for the users' loved ones. To achieve that our system will contain:

- The digital twin will be created through an input image from the user.
- The digital twin avatar will go through transformers to perform an accurate look.
- The user will have the ability to customize the avatar look.
- The user will have a conversation with the digital twin through a speech-to-text feature.
- The system will respond to the user's questions through an AI model using a provider called GPT-4 model for natural language processing is state-of-the-art.
- LLM is based on a transformer architecture and has been pre-trained on a large amount of different text data.

3.4 System Context

As shown in figure 4 this context diagram explain the connection between the core system of Digital Twin. The user entity shows that the user will be able to upload multiple images to be converted to 3D model and the final result will be shown to the user at the end by generating it. Also, voice cloning entity it's responsible of taking a record from human and convert it to be realistic to the avatar. It also shows that LLM entity is essential part of the system because it has the responsibility of generating a real human response and understanding the user input. In addition, all of the images will go through image processing stages to get the accurate structure.

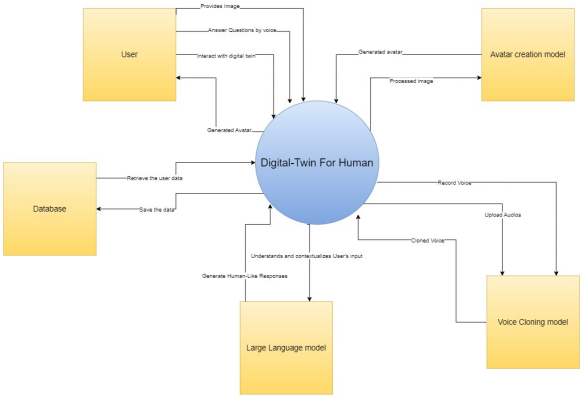

Figure 5: Context Diagram

3.5 Objectives

The main objective of our project includes the following:

- Creating a realistic digital twin from the user's input that relates to the user's physical, appearance and facial expressions to ensuring that the user have an immersive experience.
- Building an advanced NLP model to generate answers and conversation with the user while the user initials a conversation with the digital twin.
- Developing a customization system to give the ability to the user to customize the avatar which as digital-twin's appearance, voice, and other features to build an emotional connection between them.
- Integrate the TTS to enable the digital twin to achieve a conversation with natural and human-voice while responding to the user's questions.
- Developing an adaptive learning technology to the digital twin for understanding and learning from the user's preference and communication style and personality by giving the ability to the digital twin to evolve and refine itself.

3.6 User Characteristics

- Standards user are people of any age and gender who can use mobile application.
- Since English will be the user interface's default language the user must have a basic knowledge and understanding of the language.
- Digital Twin does not require any specific advanced computer knowledge to use it expect for the developers and administrator of the system.
- All the users must have the basic mobile handling skills such as connecting to the internet, and signing into the software.

4 Functional Requirements

4.1 System Functions

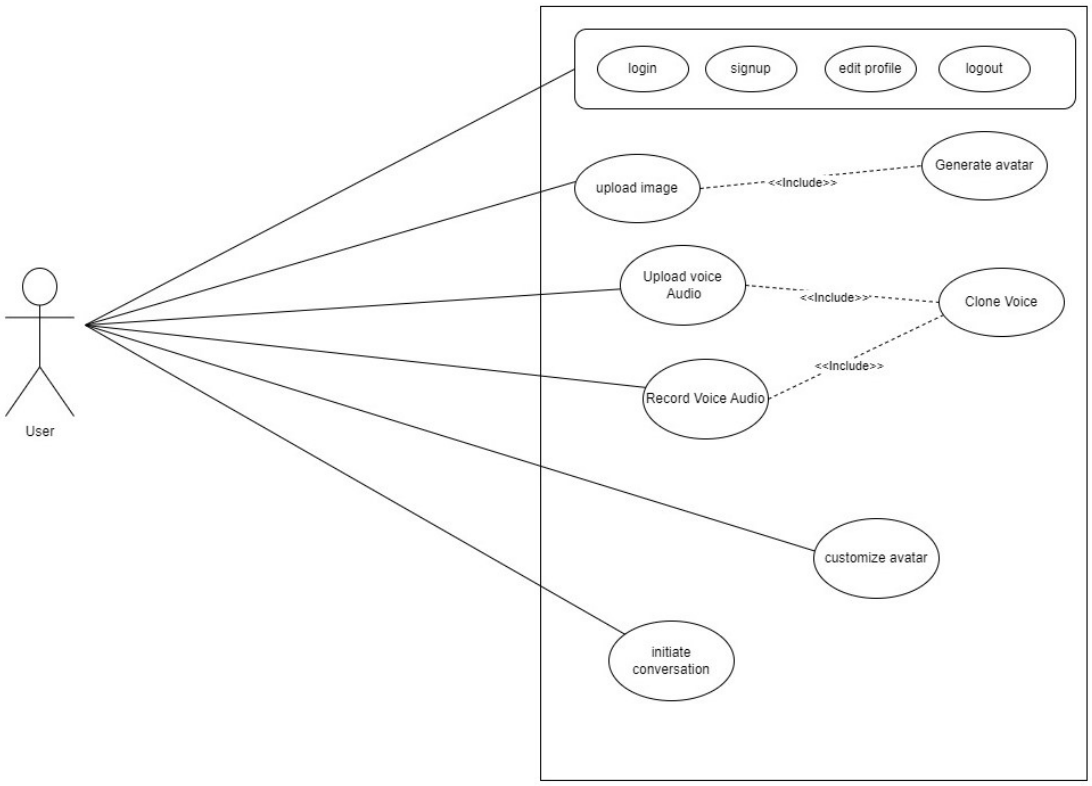

Figure 6: User Usecase Diagram

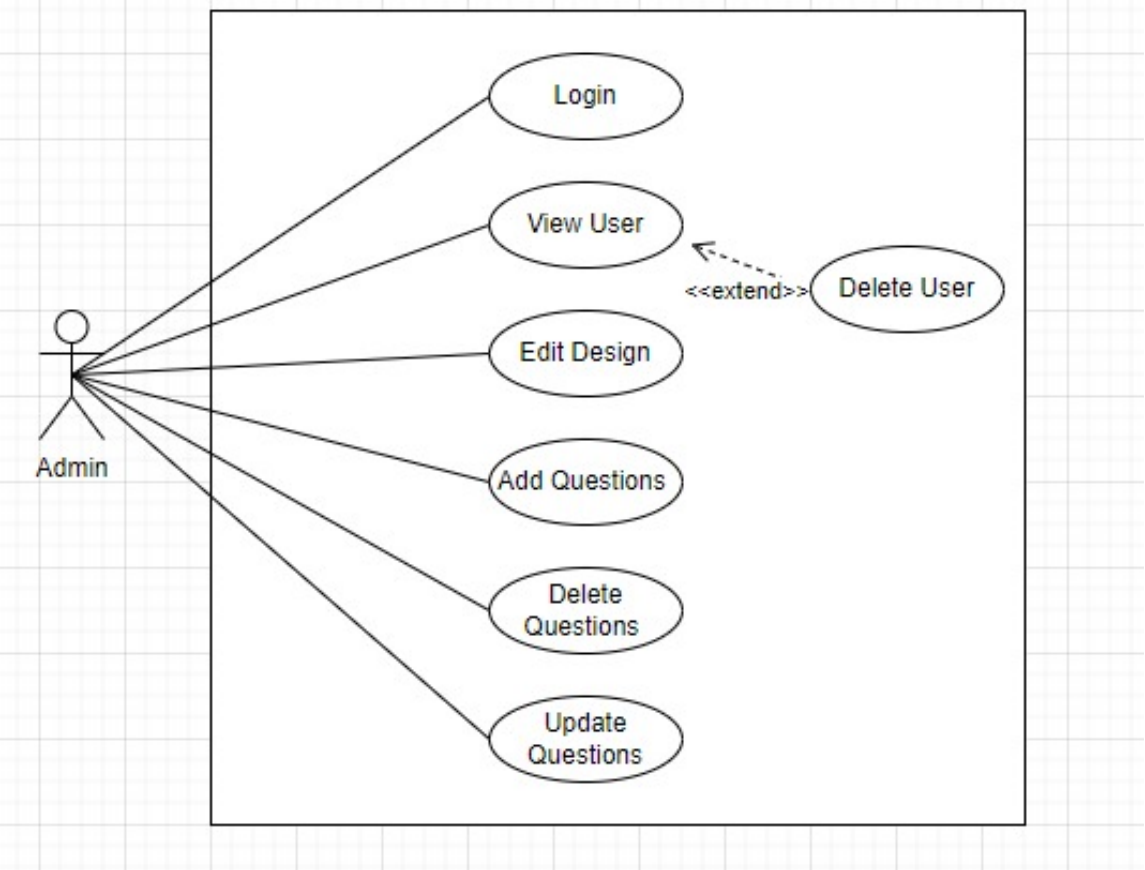

Figure 7: Admin Use-case Diagram

- **Id:01** The user shall be able to sign Up to create a new account .
- **Id:02** The user shall be able to login to their account
- **Id:03** The user shall be able to edit his profile
- **Id:04** The user shall be able to logout
- **Id:05** The user shall be able to upload images to predict a 3D Model
- **Id:06** The user should be able to upload images to show their emotions.
- **Id:07** The user shall be able answer some questions with voice
- **Id:08** The user shall be able to choose their avatar
- **Id:09** The user shall be able to customize their avatar
- **Id:10** The user shall be able to create more than one avatar
- **Id:11** the user shall be able to upload their voice
- **Id:12** The user shall be able to speak with the avatar
- **Id:13** The user shall be able to text with the avatar
- **Id:14** The system shall clone the user's voice
- **Id:15** The system shall make the avatar adapt on the user
- **Id:16** The system shall create the avatar based on user requirements
- **Id:17** The system shall make the avatar express emotions based on images and conversations.
- **Id:18** The system shall make the avatar respond on user's input by voice
- **Id:19** The system shall understand and contextualizes User's input

4.2 Detailed Functional Specification

Table 2: Conversation

Name	**chatting**
Code	Id:16
Priority	High
Critical	Cell essential for the user to communicate with the avatar
Description	After getting user's input, it start to convert the answer from text to speech so that the avatar can be able to interact with user by voice.
Input	Text, or sound command.
Output	Generated Text
Pre-condition	User must already have a created avatar
Post-condition	If created start taking input from user then start processing it
Dependency	Id:11
Risk	Low accuracy

Table 3: Voice Cloning

Name	**Cloning**
Code	Id:13
Priority	high
Critical	voice must be clear
Description	it clone user's voice, by defusing the sound the user will reply to the questions with
Input	voice records by user
Output	voice
Pre-condition	None
Post-condition	User's voice cloned
Dependency	Id:10
Risk	unclear voice

Table 4: Question

Name	**Edit account**
Code	Id:03
Priority	Medium
Critical	None
Description	It allows user to Edit the data in the database
Input	Email Password First and last name and mobile number
Output	Boolean(data updated or not)
Pre-condition	user must be logged in
Post-condition	If account Updated successfully
Dependency	Id:01
Risk	None

Table 5: 3D Model

Name	**3D Model**
Code	Id:15
Priority	High
Critical	None
Description	3D Model is generated from uploaded photos of the user by detecting the characteristics of the images
Input	at least one image to be uploaded
Output	Accurate 3D Model of the user
Pre-condition	user must upload one image
Post-condition	If Data has no errors a 3D model is viewed
Dependency	Id:05
Risk	None

Table 6: Emotion

Name	**Emotion**
Code	Id:17
Priority	high
Critical	image of facial expression must be clear
Description	After uploading an image of facial expression, the digital twin will have the same facial expression as the user.
Input	images and text
Output	the digital twin will be able to express feelings
Pre-condition	at least 3 images of facial expression
Post-condition	User's facial expression uploaded
Dependency	Id:06
Risk	None

Table 7: Customize Avatar

Name	**Customize Avatar**
Code	Id:09
Priority	high
Critical	None
Description	After the user have created the avatar and want to change in his look. the avatar will be customized upon the user needs
Input	Selecting the assets (outfits,Hair,skin)
Output	The avatar will be customized
Pre-condition	Selecting the chosen avatar
Post-condition	avatar will be customized according to the user changes
Dependency	Id:16
Risk	None

5 Design Constraints

5.1 Standards Compliance

Digital Twin Of Human can only runs on Android or IOS operating systems. Users can download it from Play Store for Android devices or App Store for iOS devices, however it is only compatible with smartphones. In order to connect to the internet. So, the mobile device needs to be a smartphone.

5.2 Hardware Limitations

The user must have a smartphone connected to the internet, with a working camera to be able to take pictures of themselves or a photo for the beloved ones from their devices to upload it to the system. Additionally, a mobile device with a working microphone and speakers to be able to record the voice. So, our application can make an avatar has the same characteristics and the same face from the pictures.

5.3 Network Constraints

For the application to function properly a stable internet connection is required.

6 Non-functional Requirements

6.1 Security

The user registration process the system will encrypt the password when the user login.

Admin will create an account for them to access the system's data and features.

6.2 Reliability

The user information will be kept in the database and is always up to date it can not lost and always be obtained from the administrator with the user's permission.

6.3 Privacy

All user's data will be secured and unavailable to anyone.

6.4 Availability

The application will be available on Play store for android users and App store for iOS users.

6.5 Usability

The system should be developed in a user-friendly.

The system should be offered in any language.

The system should be fully automated as much as possible for users.

The system should be simple and easy to understand for user.

7 Data Design

Alpaca is a dataset of 52,000 guidelines and examples generated by OpenAI's text-davinci-003 engine. By using this instruction data, language models can go through instruction-tuning, enhancing the language model's ability to follow instructions. The authors built on the data generation pipeline from Self-Instruct framework and made some modifications such as using text-davinci-003 engine instead of DaVinci, more aggressive batch decoding was used, single instance was generated for each instruction, instead of 2 to 3 instances as in Self-Instruct.

instruction string · *lengths* 9 489	**input** string · *lengths* 0 2.47k	**output** string · *lengths* 0 4.18k	**text** string · *lengths* 154 4.5k
Arrange the words in the given sentence to form a…	quickly the brown fox jumped	The quick brown fox jumped quickly.	Below is an instruction that describes a task,…
Analyze the given text for its tone.	The world has been greatly impacted by the…	The tone of the text is one of concern and…	Below is an instruction that describes a task,…
Use the given data to calculate the median.	[2, 3, 7, 8, 10]	The median of the given data is 7.	Below is an instruction that describes a task,…
Rewrite the given paragraph in a shorter,…	Although it is generally accepted that the…	The internet has allowed us to connect globally,…	Below is an instruction that describes a task,…
Design a logo for a website about recycling.		The logo should feature a green motif that is…	Below is an instruction that describes a task.…
Generate a poem that expresses joy.		Joy flows through me like a river clear, Bringing…	Below is an instruction that describes a task.…

Figure 8: A sample of Alpaca dataset

An example of "train" looks as follows:


```
{
    "instruction": "Create a classification task by clustering the given list of items.",
    "input": "Apples, oranges, bananas, strawberries, pineapples",
    "output": "Class 1: Apples, Oranges\nClass 2: Bananas, Strawberries\nClass 3: Pineapples",
    "text": "Below is an instruction that describes a task, paired with an input that provides further contex
}
```


Figure 9: Trained model

ERD Diagram:

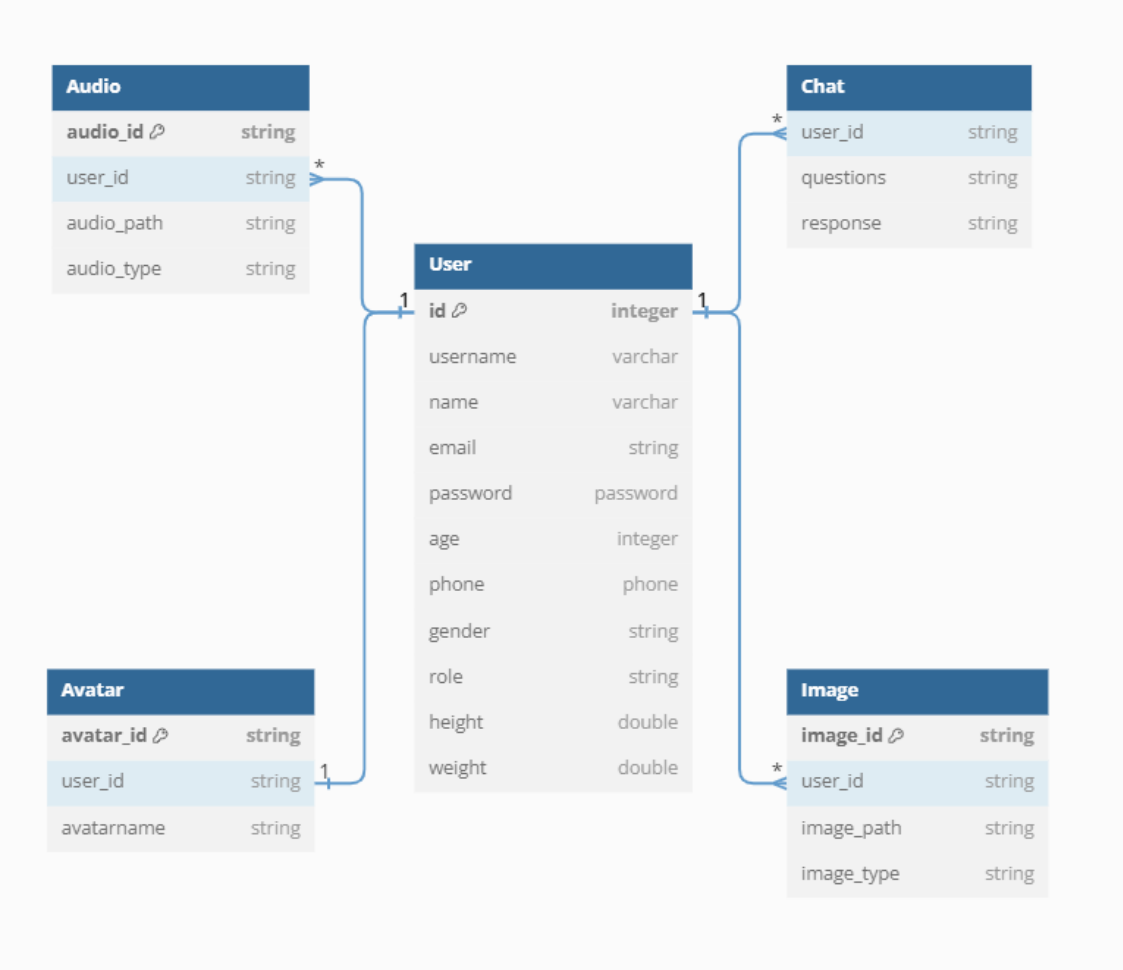

Figure 10: Entity Relationship Diagram

ERD Diagram:

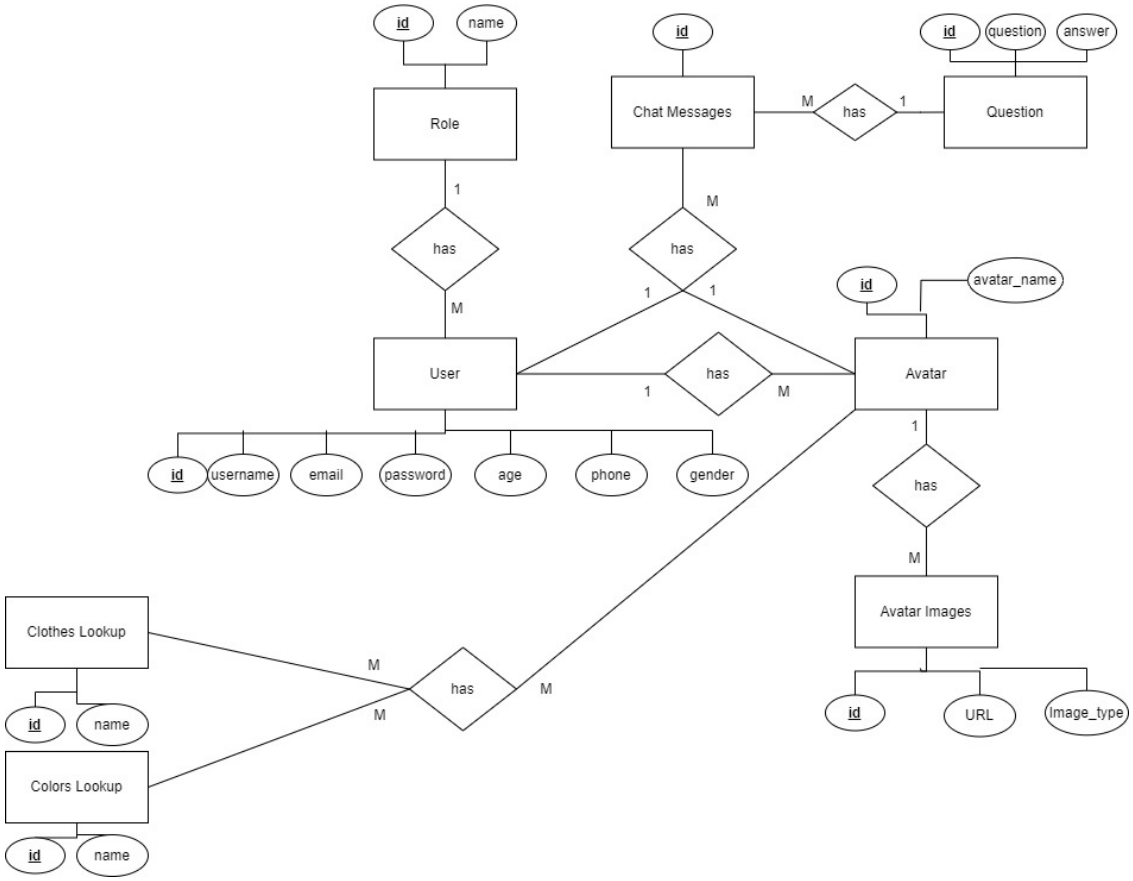

Figure 11: Entity Relationship Diagram

8 Preliminary Object-Oriented Domain Analysis

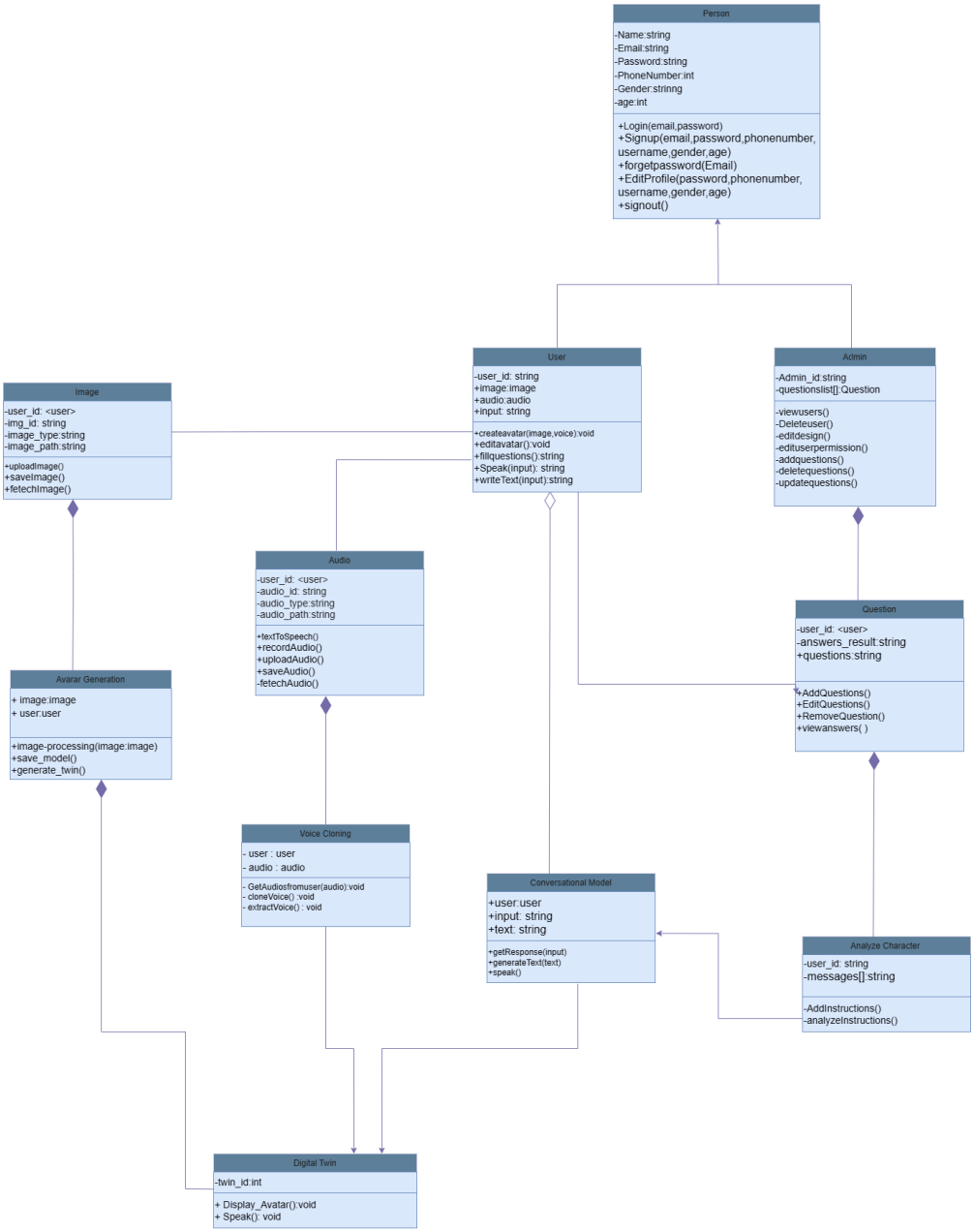

Figure 12: Class Diagram

9 Operational Scenarios

9.1 User Scenario

9.1.1 Upload Image

First the user will have to sign up. The user will upload an image, then the system will process the image and generate a lifelike 3D digital twin avatar based on the image.

9.1.2 Chat With Avatar

The user can interact with the avatar by texting or voice chatting giving commands or questions.

9.1.3 Record Voice

The user will record their voice to make their avatar have the same voice.

9.1.4 Upload Voice

The user can upload a file of their voice to make their avatar have the same voice or to their beloved ones.

9.1.5 Voice Cloning

Once the avatar has been generated. The user will have to answer some questions to analyze their personality while this operation our system will record his voice to use it in the voice cloning.

9.1.6 Fill Questions

The user has to answer some question about their life, attitude or a specific situation and how he/she will handle that situation and also they will record their facial expression to make their avatar has the same concept as thinking.

9.1.7 View Avatar

The user after generates the avatar and have his voice, the user will explore the avatar by engages in virtual conversation and immersive interaction with the avatar.

9.2 Admin Scenario

9.2.1 View Users

The admin will log in by entering the username and password and then, find all user's avatars and how active each user is. Also he/she can edit, delete user from the application.

9.2.2 View Questions

The admin has an option to add, edit and delete questions for the user. also he can edit user permission.

10 Project Plan

Task	Start date	End date	Duration	Member
Ideas and Supervisor	25/08/2023	30/09/2023	35 Days	All Members
Information collection and research	3/10/2023	21/10/2023	18 Days	All Members
Survey and proposal preparation	6/11/2023	15/11/2023	9 Days	All Members
Proposal Presentation 10%	16/11/2023	19/11/2023	3 Days	All Members
Classify dataset	21/11/2023	5/12/2023	21 Days	All Members
Information collection and research	06/12/2023	10/12/2023	3 Days	All Members
Searching for bigger language models	11/12/2023	16/12/2023	5 Days	All Members
Searching for bigger dataset	16/12/2023	19/12/2023	3 Days	All Members
Code preparation	21/12/2023	27/12/2023	6 Days	All Members
SRS Preparation	28/12/2023	14/1/2024	17 Days	All Members
SRS Presentation 35%	15/1/2024	17/1/2024	2 Days	All Members
SDD Preparation	20/1/2024	4/2/2024	15 Days	All Members
SDD Presentation 65%	5/3/2024	8/3/2024	3 Days	All Members
Mobile Application Development	10/3/2024	15/4/2024	36 Days	All Members
Platform Development	20/3/2023	10/4/2024	21 Days	All Members
Prototype Submission 80%	25/3/2024	25/4/2024	33 Days	All Members
Technical evaluation 90%	1/5/2024	7/5/2024	7 Days	All Members
Testing and Validation	1/5/2024	15/5/2024	15 Days	All Members
Final Thesis 100%	20/5/2024	15/6/2024	25 Days	All Members

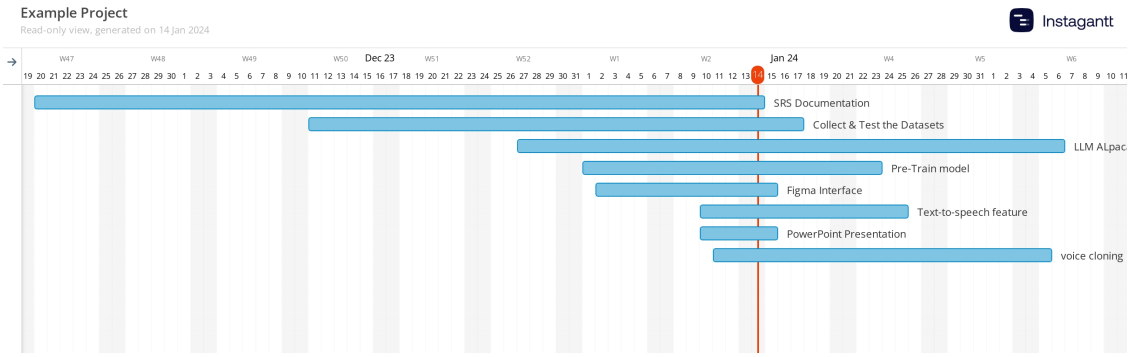

Figure 13: GANTT Chart

11 Appendices

11.1 Definitions, Acronyms, Abbreviations

Table 8: Definitions

Abbreviations	Definitions
NLP	Natural Language processing
TTS	Text-to-speech
STT	Speech-to-text
SMPL	Simple Modeling and Processing Language
LLM	Large Language model

11.2 Supportive Documents

11.2.1 Dataset

- Alpaca is a dataset of 52,000 guidelines and examples generated by OpenAI's text-davinci-003 engine. By using this instruction data, language models can go through instruction-tuning, enhancing the language model's ability to follow instructions. The authors built on the data generation pipeline from Self-Instruct framework and made some modifications such as using text-davinci-003 engine instead of davinci, more aggressive batch decoding was used, single instance was generated for each instruction, instead of 2 to 3 instances as in Self-Instruct.

- we are going to use the SMPL dataset to ptovide accurate representation of the human body and also we will use another datasets (AIST++, SHHQ and DEEPFashion) to test the quality of the generated modules. These datasets are mainly contaion of, AIST++ is a large dataset consisting of 10.1M images covering 30 different performers in 9 camera views. DEEP-FASHION and SHHQ are single-view image datasets consisting of 8k and 40k identities.

11.2.2 users/survey

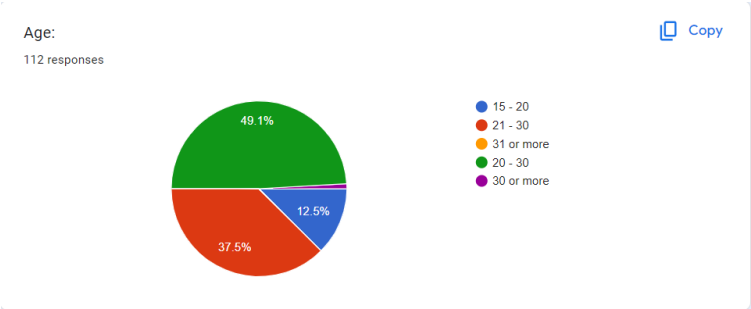

Figure 14:

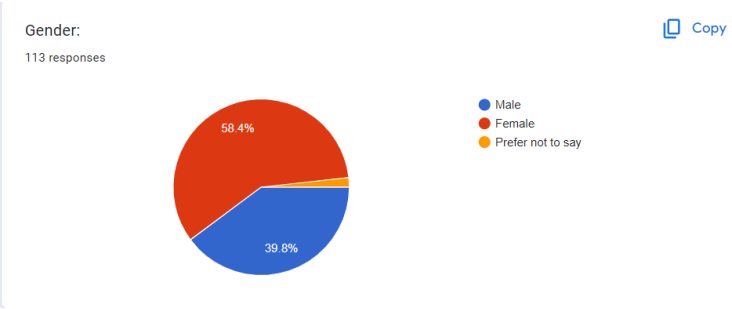

Figure 15:

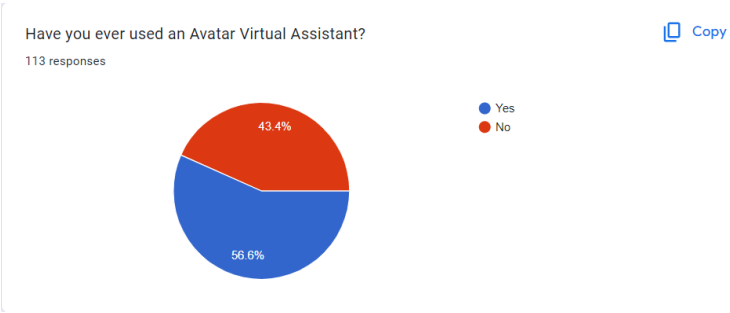

Figure 16:

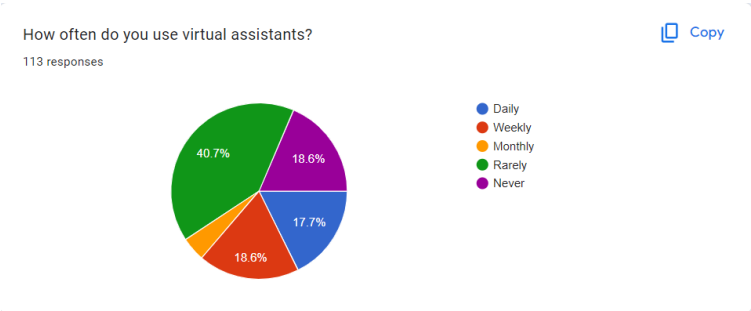

Figure 17:

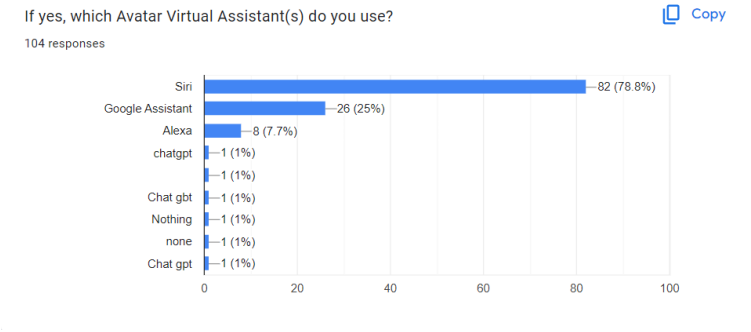

Figure 18:

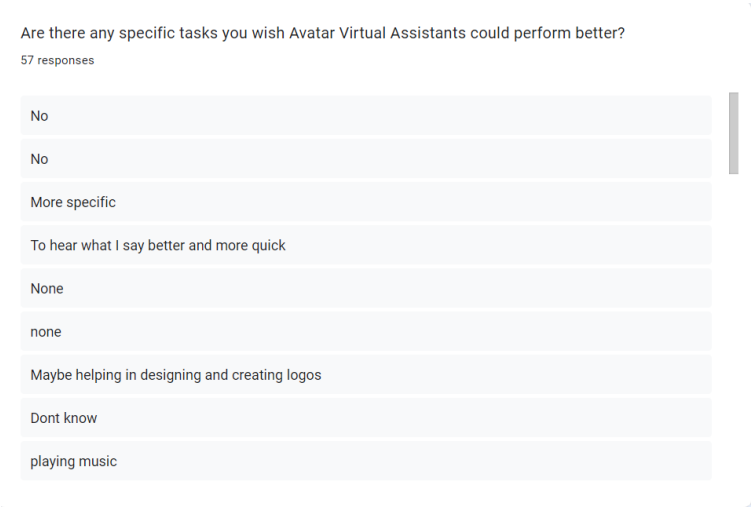

Figure 19:

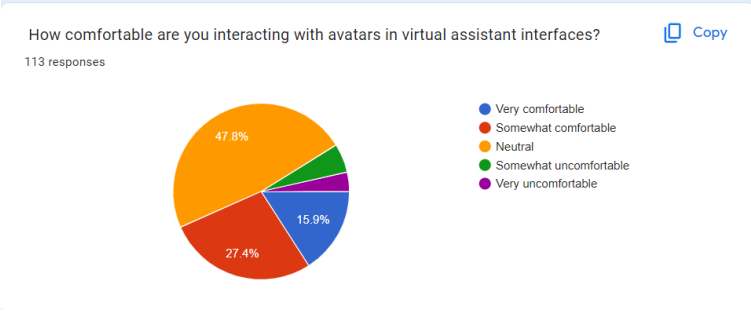

Figure 20:

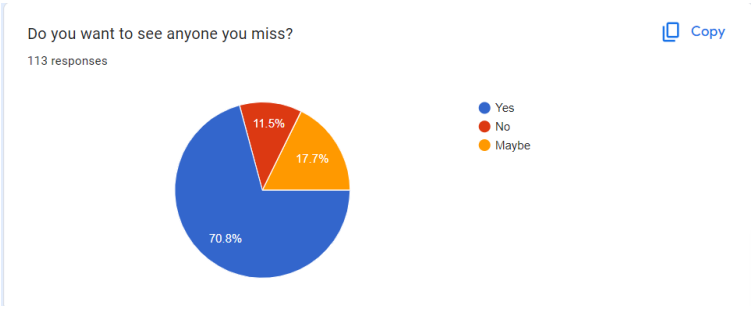

Figure 21:

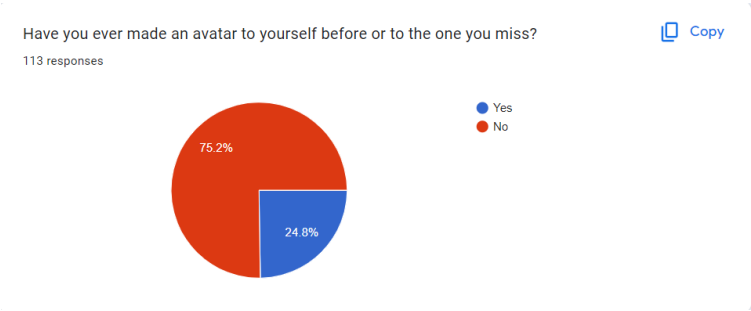

Figure 22:

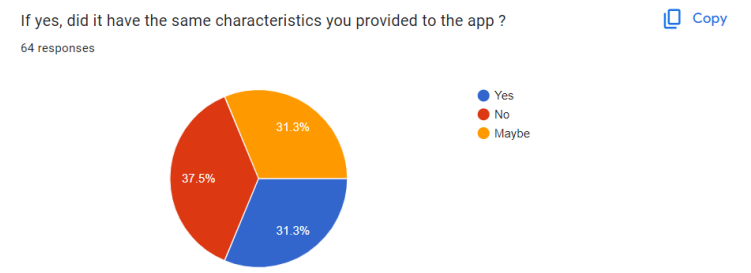

Figure 23:

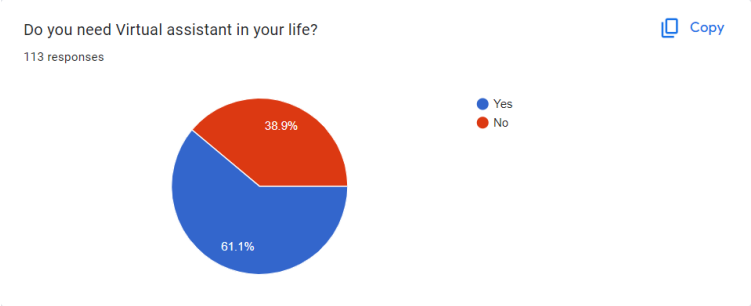

Figure 24:

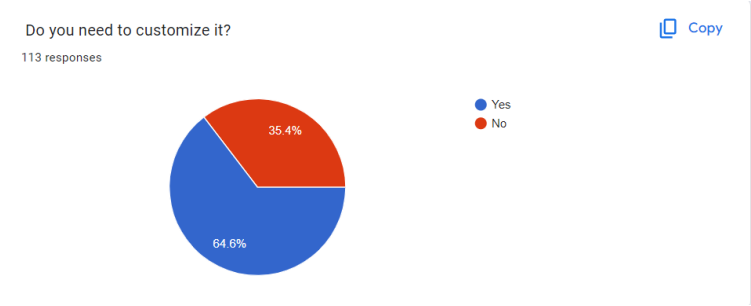

Figure 25:

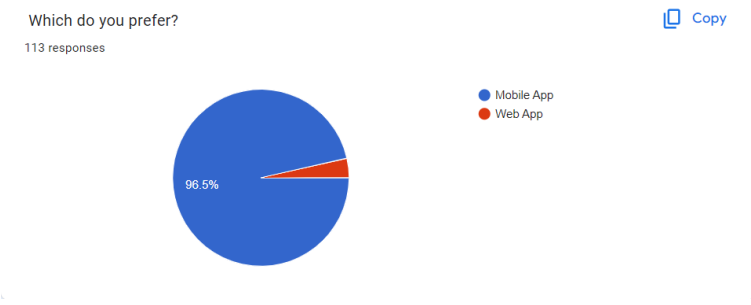

Figure 26:

Figure 27:

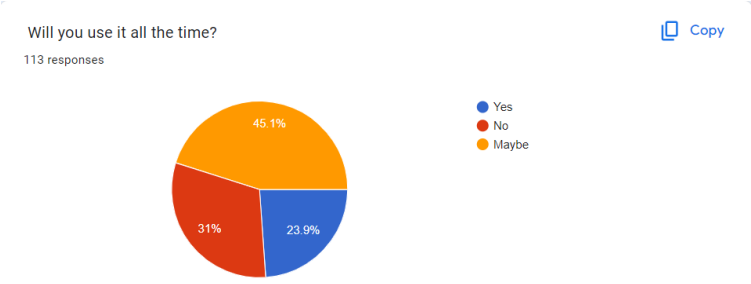

Figure 28:

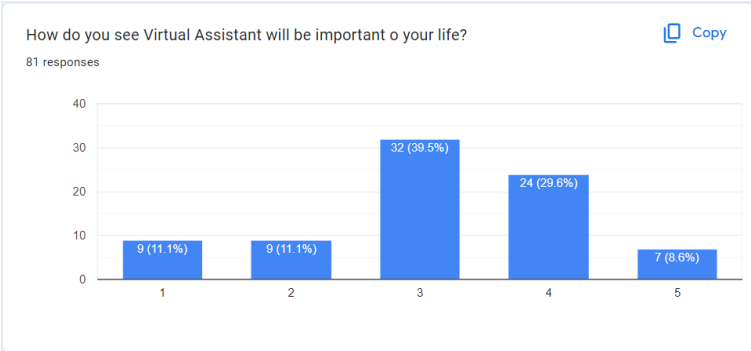

Figure 29:

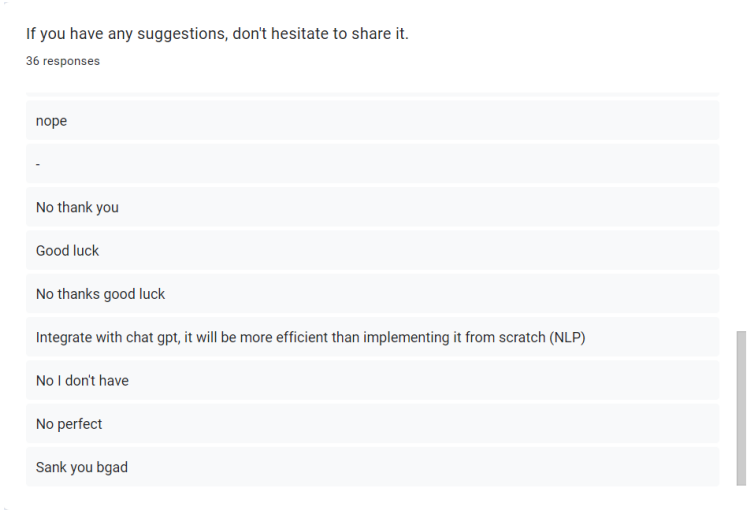
If you have any suggestions, don't hesitate to share it.
36 responses
nope
-
No thank you
Good luck
No thanks good luck
Integrate with chat gpt, it will be more efficient than implementing it from scratch (NLP)
No I don't have
No perfect
Sank you bgad

Figure 30:

References

[1] Andrej Miklosik, Nina Evans, and Athar Mahmood Ahmed Qureshi. “The use of chatbots in digital business transformation: A systematic literature review”. In: *IEEE Access* 9 (2021), pp. 106530–106539.

[2] Tim Dettmers, Artidoro Pagnoni, Ari Holtzman, et al. “QLORA: Efficient Finetuning of Quantized LLMs”. In: (May 2023). Preprint available on arXiv. arXiv: arxiv.org/pdf/2305.14314.pdf. URL: https://arxiv.org/abs/2305.14314.

[3] Ashish Vaswani, Noam Shazeer, Niki Parmar, et al. “Attention Is All You Need”. In: (June 2017). Published as a conference paper in Advances in Neural Information Processing Systems (NeurIPS). arXiv: arxiv.org/pdf/1706.03762.pdf. URL: https://arxiv.org/abs/1706.03762.

[4] Wayne Xin Zhao, Kun Zhou, Junyi Li, et al. “A survey of large language models”. In: *arXiv preprint arXiv:2303.18223* (2023).

[5] Replika. *Replika Website*. URL: https://replika.com/.

[6] *Create Full-Body 3D Avatar from a Photo | Ready Player Me*. URL: https://readyplayer.me/hub.