TRIBHUVAN UNIVERSITY
INSTITUTE OF ENGINEERING

Khwopa College of Engineering
Libali, Bhaktapur
Department of Computer Engineering

A FINAL REPORT ON

AvatarFusion: 3D Companion with Sentiment Analysis

Submitted in partial fulfillment of the requirements for the degree

BACHELOR OF COMPUTER ENGINEERING

Submitted by

Bishal Bhatt	KCE076BCT015
Indira Kasichhwa	KCE076BCT017
Niru Dhaubanjar	KCE076BCT021
Rahul Khatri	KCE076BCT029

Under the Supervision of
Er. Prakash Chandra Prasad
Department of Electronics and Computer Engineering

Khwopa College Of Engineering
Libali, Bhaktapur
March 11, 2024

CERTIFICATE OF APPROVAL

This is to certify that this major project entitled **"AvatarFusion: 3D Companion with Sentiment Analysis"** submitted by Bishal Bhatt(KCE076BCT015), Indira Kasichhwa (KCE076BCT017), Niru Dhaubanjar(KCE076BCT021), Rahul Khatri(KCE076BCT029) has been examined and accepted as the partial fulfillment of the requirements for degree of Bachelor in Computer Engineering.

.....
Er. Bibha Sthapit
External Examiner
Assistant Professor
Dept. of Electronics and Computer Engineering
IOE, Pulchowk

.....
Er. Prakash Chandra Prasad
Project Supervisor
Assistant Professor,
Department of Electronic and Computer Engineering,
IOE, Pulchowk

.....
Er. Dinesh Gothe
Head of Department,
Department of Computer Engineering
Khwopa College of Engineering

Copyright

Acknowledgement

We would like to express our heartfelt gratitude to Khwopa College of Engineering for providing us with this wonderful opportunity to work on this project. Our sincere thanks go to our supervisor, Er.Prakash Chandra Prasad, for his invaluable guidance, encouragement, and unwavering support throughout the course of this project. We would also like to extend our gratitude to Khwopa College of Engineering for providing us with the necessary resources and facilities to complete this project.

We are deeply indebted to our Head of Department, Er. Dinesh Gothe, for his constant motivation, advice, and support during our Bachelor's program. We are grateful for his willingness to share his knowledge and expertise with us, which has been instrumental in shaping our career.

We would like to acknowledge the contributions of our classmates who provided us with their valuable suggestions, which were useful in various phases of the project. Their support and collaboration have been invaluable, and we are fortunate to have had such a dedicated and talented group of individuals to work with. Finally, we would like to extend our appreciation to everyone who played a role, directly or indirectly, in making this project a success.

Bishal Bhatt	KCE076BCT015
Indira Kasichhwa	KCE076BCT017
Niru Dhaubanjar	KCE076BCT021
Rahul Khatri	KCE076BCT029

Abstract

This project centers on developing an advanced 3D avatar chatbot aimed at redefining the virtual communication experience. To tackle observed limitations in current projects, such as static avatars and limited conversational capabilities, the focus is on resolving these shortcomings. Traditional chatbots lack a genuine sense of presence and personalization, impeding effective communication. The project integrates advanced 3D modeling and animation techniques to craft realistic avatars with expressive gestures and fluid movements, enhancing overall engagement. Furthermore, incorporating advanced natural language processing algorithms enables the chatbot to comprehend and respond to user input with unparalleled precision. The project offers numerous benefits, including heightened user engagement, enriched chat experiences, and improved accessibility for individuals with diverse communication preferences. In essence, this 3D avatar chatbot project marks a significant advancement in virtual communication, surpassing the constraints of existing models and furnishing a more interactive and authentic platform for users across various domains. The final assessment underscores its potential to transform digital communications, offering a glimpse into the future of engaging and personalized virtual interactions.

Keywords: *Animation, Chatbot, Dynamic Interaction, Virtual communication, 3D modeling*

Contents

List of Figures

List of Tables

List of Abbreviation

Abbreviations	**Meaning**
Adam	Adaptive Moment Estimation
API	Application Programming Interface
BOW	Bag Of Words
BERT	Bi-directional Encoder Representation from Transformers
BLEU	Bilingual Evaluation Understudy
BPTT	Backpropagation Through Time
CBT	Cognitive Behaviour Therapy
DDPDM	Data-Driven Persona Development Method
GPT	Generative Pre-trained Transformer
KDD	Knowledge discovery in databases
LSTM	Long Short Term Memory
NLP	Natural Language Processing
RNN	Recurrent Neural Network
ROC-AUC	Receiver Operating Characteristic - Area Under the Curve
Seq2Seq	Sequence to Sequence
SER	Sign Error Rate
STT	Speech To Text
TER	Text Emotion Recognition
TF-IDF	Term Frequency - Inverse Document Frequency
TTS	Text-To-Speech
UEQ	User Experience Questionnaire

Chapter 1

Introduction

1.1 Background Introduction

Conversational agents or chatbots are text-based dialogue systems integrated in mobile apps or web pages. The chatbot simulates a realistic conversation partner by giving the user appropriate written answers in a language that he or she understands [1]. These chatbots can be used in various contexts, such as customer service, information acquisition, educational support, and entertainment. Conversational chatbots have become an increasingly popular tool in recent years. The software applications are created to imitate dialogues with individuals by utilizing NLP techniques such as ChatGPT [2] and applying them to a variety of applications, including casual entertainment purposes, customer services , and educational purposes to assist teachers and students .

Ranging from casual and open-domain to more domain-specific and fact-based, these chatbots are built using various deep learning models, such as RNN , Seq2Seq , LSTM, BERT [3], GPT , as well as leveraging different training techniques, such as reinforcement learning or transfer learning, in order to improve the performance of NLP algorithms and chatbots. A recent popular, encouraging example is the success of ChatGPT, which received a great amount of attention and inspired researchers to generate more ideas regarding chatbot applications [4]. A review of Laranjo et al. confirms that the use of conversational agents with unconstrained natural language input capabilities for health-related purposes is an emerging field of research [5].A chatbot, powered by AI and NLP, acts as a virtual assistant capable of engaging in human-like conversations. By integrating a chatbot into your project, you unlock a multitude of benefits and opportunities that can significantly enhance its overall performance and user experience.Many chatbots have been developed for providing mental health interventions. For example, the chatbot "Wysa" uses several evidence-based therapies (e.g. cognitive behavioural therapy, behavioural reinforcement, and mindfulness) to target symptoms of depression for users. LISSA is another chatbot that provides training for people with autism in order to develop their social skills [6]. The majority of chatbots (92.5%) depended on only decision trees to generate their responses; only 7.5% used machine learning approaches.

The text interface used in chatbots is not always entirely efficient or engaging. To enhance the efficiency of chatbots, one solution is to implement interfaces apart from text. One such example is the use of a voice interface combined with a 3D avatar.Avatar which is able to display emotions and moods by giving face expression on conversation that been talked so the subject becomes more interesting. With this chatbot, the interaction between the bot and user becomes more evident just because of the facial expression [7].

1.2 Motivation

We identified the limitations of text-based chatbots in terms of efficiency and lack of emotional expression [1]. Internet chatbots typically rely on text for input and output. While text-based input and output have their advantages, such as allowing users to review their input for errors, they are not efficient due to the reliance on a keyboard. Moreover, the lack of visual representation limits the display of emotions in conversations, making the text interface unattractive [7]. As a result, chatbots are not fully efficient and lack appeal.

The creation of a virtual companion with a visually appealing 3D avatar is driven by the desire to revolutionize user experience in virtual communication. Traditional projects often fall short in providing realistic and dynamic interactions due to static avatars and limited conversational capabilities. In this context, our project seeks to address these limitations by leveraging advanced 3D modeling and animation techniques, enhancing avatars with expressive expressions, gestures, and fluid movements for a more engaging experience. This endeavor not only addresses the shortcomings observed in current projects but also aspires to set a new standard in the realm of virtual communication.

1.3 Problem Definition

By introducing a 3D avatar and sentiment analysis, our aim is to enhance human-machine interaction and create a more engaging conversation experience. In addition to conversation, 3D avatars can enhance the attractiveness of chatbots and effectively display emotions through facial expressions, making the conversation more engaging. Facial expressions play a crucial role in non-verbal communication, without the need for words. Facial expressions are not exclusive to humans but also observed in animals, where they serve as a means of defense or offense. For example, a dog may exhibit an angry face while barking to deter potential threats without physical contact. This chatbot enhances interaction through facial expressions. Research suggests that interactive animated characters can benefit individuals with social difficulties. The chatbots instructions become clearer, and users feel like they are interacting with a human rather than a bot.

1.4 Goals and Objectives

The main objective of this project is:

- To create a virtual companion with visually appealing 3D avatar that interacts and responds to the users.

1.5 Scope and Applications

The scope of our project includes:

- Develop a 3D avatar chatbot web application..
- Adhere to user-centered design principles for a personalized user experience.
- Integrate emotional engagement and support features.

Chapter 2

Literature Review

AvatarFusion, a 3D companion with sentiment analysis, is an innovative technology that combines virtual avatars and emotion recognition algorithms to create interactive and emotionally responsive digital companions.Numerous articles exploring AI in conversational contexts and analyzing existing chatbots have been published. However, the utilization of artificial intelligence for emotional well-being and Visually appealing chatbot is still in its nascent phase of development. Below are some of the papers related to this field.

Ekaterina Svikhnushina,et.al ,at 2021.conducted evaluation to validate the model's constructs and establish meaning- ful causal relationship and proposed paper [8].The paper [8] presents a large-scale survey aimed at capturing users' preferences, expectations, and concerns regarding conversational chatbots. The authors utilized a consolidated model to gather data from a significant number of participants. Additionally, they employed structural equation modeling methods to validate and analyze the collected data. By combining these approaches, the study provides valuable insights into user perspectives on conversational chatbots, contributing to a better understanding of user needs and guiding the development of more effective chatbot systems. The outcome supports the consistency, validity, and reliability of the model, which authors called PEACE (Politeness, Entertainment, Attentive Curiosity, and Empathy).

P Antonius Angga, W Edwin Fachri, A Elevanita1, Suryadi, R Dewi Agushinta,et. al. at 2015. [7] proposed artiucle for chatbot with liveliness conversation. This paper proposes a design for a chatbot with avatar and voice interaction to enhance the naturalness and liveliness of conversations. The approach involves utilizing multiple APIs to enable speech recognition, text-based chatbot responses, and text-to-speech synthesis. The computer renders an avatar synchronized with the audio reply. The design has potential applications in customer service and other human interactions, and it can be extended by incorporating additional tools such as webcam analysis of user emotion and reactions.

Yu-Ting Wan1, Cheng-Chun Chiu, Kai-Wen Liang, Pao-Chi Chang ,et. al. at 2019. [9] presents a novel model of a chat robot that incorporates facial expression interaction and role-playing to enhance user experience. A text-sentiment-analysis network is integrated to recognize the emotion of response sentences. Additionally, a 3D-constructed anime character named "Midoriko" is introduced to enhance the visual representation of the chat robot. The technical aspects of this model rely on neural network techniques.

Debashis Das Chakladar, Pradeep Kumar, Shubham Mandal, Partha Pratim Roy, Masakazu Iwamura, Byung-Gyu Kim ,et. al. at 2021. [10] proposed the sys-

tem. In this study, a 3D avatar-based sign language learning system for Indian Sign Language (ISL) is developed. The system converts input speech/text into corresponding sign movements using NLP. It consists of three modules: speech-to-English conversion, English-to-ISL conversion, and avatar motion definition. The translation module achieves a 10.50 SER score, demonstrating its effectiveness in facilitating communication with hearing-impaired individuals.

Kerstin Denecke, Sayan Vaaheesan, Aaganya Arulnathan,et. al. at 2021. [1] proposed mobile application for mental health.In their paper, the authors introduce SERMO, a mobile application with an integrated chatbot that incorporates CBT techniques to provide support for individuals with mental health conditions. The application prompts users daily to report on events and emotions, automatically determining the basic emotion through natural language processing. SERMO offers appropriate interventions such as activities or mindfulness exercises based on the identified emotion. Additional features include an emotion diary, a list of pleasant activities, mindfulness exercises, and information on emotions and CBT. User experience evaluation with 21 participants using the UEQ revealed positive ratings for efficiency, perspicuity, and attractiveness, while evaluations related to hedonic quality showed neutral responses.

Simone Borsci, Simone Borsci, Alessio Malizia,et. al. at 2021. [11] did survery for chatbot attributes. This work conducts four studies to define attributes and develop standardized tools for assessing chatbot interaction and user satisfaction. The tools include a checklist (BOT-Check) and a questionnaire (BOT Usability Scale, BUS-15) with good psychometric properties, providing a reliable way to evaluate chatbot experiences and consider a broader range of aspects. Further testing and validation are needed for the questionnaire.

Fatima Ali Amer Jid Almahri, David Bell, Mahir Arzoky, et. al. at 2021. [12] introduces a data-driven persona development .This research utilizes design science research in three iterations to enhance student engagement using chatbots. This chapter focuses on personas elicitation, proposing a DDPDM using k-means clustering. Two datasets are analyzed, resulting in eight personas and contributing to the literature with the DDPDM and persona template for university students. Future work includes the remaining iterations.

Table 2.1: Literature Review Matrix

S.N	Title	Author	Date	Conclusion
1	Key Qualities of Conversational Chatbots – the PEACE Model	Ekaterina Svikhnushina [8]	2021	They conducted a psychometric evaluation to validate the model's constructs and establish meaningful causal relationships. The study expanded on existing criteria and provided a comprehensive understanding of user requirements for socially-aware chatbots. The findings offer design implications for future development of emotionally and socially aware conversational agents.
2	Design of chatbot with 3D avatar, voice interface, and facial expression [7]	P Antonius Angga, W Edwin Fachri, A Elevanita1, Suryadi, R Dewi Agushinta	2015	The article proposes a design for a chatbot that incorporates avatar and voice interaction to enhance the liveliness of conversations. However, the design is limited to text-to-speech recognition, generating an audio version of the chatbot's reply. The chatbot's response is then synchronized with an avatar's gestures and lip movements rendered by the computer.
3	Midoriko Chatbot: LSTM-Based Emotional 3D Avatar [9]	Yu-Ting Wan1, Cheng-Chun Chiu, Kai-Wen Liang, Pao-Chi Chang	2019	In this model, a text-sentiment-analysis network is integrated into a chatbot with facial expression interaction. It recognizes the emotion of response sentences and incorporates a 3D-constructed anime character named "Midoriko" to enhance the visual representation of the chatbot.
4	3D Avatar Approach for Continuous Sign Movement Using Speech/Text [10]	Debashis Das Chakladar, Pradeep Kumar, Shubham Mandal, Partha Pratim Roy, Masakazu Iwamura, Byung-Gyu Kim	2021	The system described in this work is a 3D avatar-based sign language learning system that utilizes Natural Language Processing (NLP) techniques to convert input speech/text into corresponding sign movements for International Sign Language (ISL).

5	A Mental Health Chatbot for Regulating Emotions (SERMO) - Concept and Usability Test [1]	Kerstin Denecke, Sayan Vaaheesan, Aaganya Arulnathan	2021	They introduced SERMO, a mobile application that combines a chatbot with methods from cognitive behavior therapy (CBT). The application aims to support individuals with mental health concerns in managing emotions and addressing thoughts and feelings.
6	The Chatbot Usability Scale: the Design and Pilot of a Usability Scale for Interaction with AI-Based Conversational Agents [11]	Simone Borsci, Simone Borsci, Alessio Malizia	2021	This work comprises four studies (literature review, survey, focus groups, testing) involving 141 participants to define chatbot interaction attributes and design a satisfaction scale.
7	Applications of Machine Learning in Education: Personas Design for Chatbots [12]	Fatima Ali Amer Jid Almahri, David Bell, Mahir Arzoky	2021	This introduces a data-driven persona development method (DDPDM) that utilizes machine learning, specifically k-means clustering, for personas elicitation in the first iteration.

Chapter 3

Requirement Analysis

3.1 Software Requirement

Software requirement for the prepared system includes:

1. Python
2. Star UML
3. Visual Studio Code
4. Google Colaboratory
5. CUDA
6. Git
7. React Three fiber
8. Texmaker
9. Beamer
10. Trello
11. Microsoft Team
12. Google Drive
13. Blender

3.2 Hardware requirements

3.2.1 Graphical Processing Unit

To train our model, we utilized the GPU resources offered by Google Colab for our completed projects.

3.2.2 Computer

To train, configure, and operate our system, we utilized a laptop or desktop computer with a minimum Intel-i5 processor for our completed projects.

3.2.3 Microphone

A microphone was required to capture speech input from the user for our completed projects. Typically, we utilized the computer's built-in microphone.

3.3 Functional Requirement

3.3.1 Speech Recognition

The chatbot's requirement for speech recognition demands robust technology to accurately interpret user input. This feature ensures seamless communication by capturing natural language, allowing users to interact with the system effortlessly. The effectiveness of the chatbot heavily relies on precise speech recognition, laying the foundation for an immersive and user-friendly experience.

3.3.2 3D Avatar Rendering

This capability is fundamental to creating a lifelike virtual representation of the user, accurately reflecting their appearance and movements. The realism of the 3D avatar enhances the overall user experience, making interactions more engaging and personalized. This feature ensures that the avatar serves as an effective visual counterpart, contributing to the success of the chatbot in various applications, including virtual reality, gaming, and social platforms.

3.3.3 Avatar and Voice Synchronization

The synchronization of gestures and facial expressions with audio responses is a crucial functional requirement. This ensures that the 3D avatar dynamically mirrors the conversation, creating a cohesive and immersive experience. The seamless coordination between the avatar's movements and the spoken words enhances the realism and effectiveness of the virtual interaction, contributing to a more engaging and lifelike communication process. This synchronization is fundamental for bringing the chatbot's responses to life and making the user experience more visually compelling and interactive.

3.3.4 Natural Language Processing and Conversation Management

Avatarfusion should employ advanced NLP algorithms to understand and interpret user messages accurately. It should respond to user queries, statements, and emotions with empathy, sensitivity, and appropriate expressions. The system should have pre-defined conversation flows to guide users and provide relevant emotional support.

3.3.5 Text-to-speech conversion

Text-to-speech conversion is essential for transforming text-based responses into audio versions. This feature adds a crucial layer to the chatbot's capabilities, allowing for a more immersive and natural interaction with users. By converting textual information into spoken words, the system enhances accessibility and user engagement, contributing to a seamless and dynamic conversation. This functionality ensures that the chatbot not only generates meaningful responses but also

delivers them in an audible format, enriching the overall user experience with a human-like element

3.4 Non-Functional Requirement

These prerequisites are not required by the system, but they are necessary for the system to work at its best. The following sections concentrate on the system's non-functional requirements:

3.4.1 Security

Security is the most important property for any system. In the near future, we will add user login system so that users can save their recordings and also this will ensure privacy. By applying various techniques, we will keep the data safe and hence our system will be secured.

3.4.2 Reliability

In terms of reliability, our system needed to be robust and dependable to handle various test scenarios. In our completed projects, we endeavored to enhance the reliability of our system by conducting various tests and implementing measures to ensure its robustness.

3.4.3 Maintainability

The system was developed by modularizing and dividing it into smaller submodules for our completed projects. This approach ensured that the system remained maintainable, making it easy to inspect and manage specific submodules.

3.4.4 Portability

Portability of a system pertains to its capability to be used in different environments without impacting performance. For our completed projects, the system we developed is portable across various operating systems such as Windows and Linux.

3.4.5 Performance

The performance of our system, in our completed projects, was assessed based on the time taken to produce output once the input was provided. Additionally, we aimed to enhance the accuracy of the system by implementing state-of-the-art algorithms.

3.4.6 Usability

The system was designed to be efficient, effective, engaging, error-tolerant, and easy to learn for our completed projects. It featured simple and intuitive user interfaces, enabling the creation of an intelligent companion with emotions.

3.4.7 Scalability

The system demonstrated scalability in our completed projects, continuing to meet all its objectives and performing well as new requirements and demands increased. Our initial plan was to develop a simple application and subsequently scale it into a mobile app and a larger, more complex system.

Chapter 4

Feasibility Study

4.1 Technical Feasibility

All necessary hardware components, including a computer equipped with a robust GPU and 32GB of RAM, were readily available. Additionally, certain test programs were conducted using Google Colaboratory, ensuring no hardware-related obstacles were encountered. The development process leveraged Python, a language familiar to all team members, while other essential utility software was both freely accessible and user-friendly. Consequently, no issues arose regarding language proficiency or software compatibility. Furthermore, datasets crucial for training both the TER and Generative models were meticulously curated, filtered, and processed from online sources, ensuring their suitability for the project's needs. With hardware, software, and datasets seamlessly integrated, the project demonstrates robust technical feasibility.

4.2 Operational Feasibility

In terms of operational feasibility, the ultimate system is optimized for standard central processing units (CPUs). This system boasts an interactive nature, featuring a user-friendly interface that is effortlessly navigable and simple to utilize.

4.3 Economic Feasibility

The development of the system utilized entirely cost-free software, while the necessary hardware resources were readily accessible for model training. Deployment of the system proves seamless even on low-end personal computers, thereby minimizing deployment costs significantly. Furthermore, the scalability of the project extends to a business-level application, ensuring its adaptability and potential for expansion. Consequently, the project stands economically viable, as there are no apparent hindrances stemming from economic constraints.

4.4 Time Feasibility

According to the project timeline outlined in the Gantt chart, the project was slated for completion within a one-year timeframe. Remarkably, the project adhered to this schedule, successfully concluding within the stipulated period.

Chapter 5

Methodology

5.1 Software Development Model

5.1.1 Agile methodology

Agile is an iterative approach to project management and software development that employ continual planning, learning, improvement, team collaboration, evolutionary development, and early delivery. It encourages flexible responses to change.

Scrum is an Agile-based project management framework where teams develop products in short project cycles called sprints.

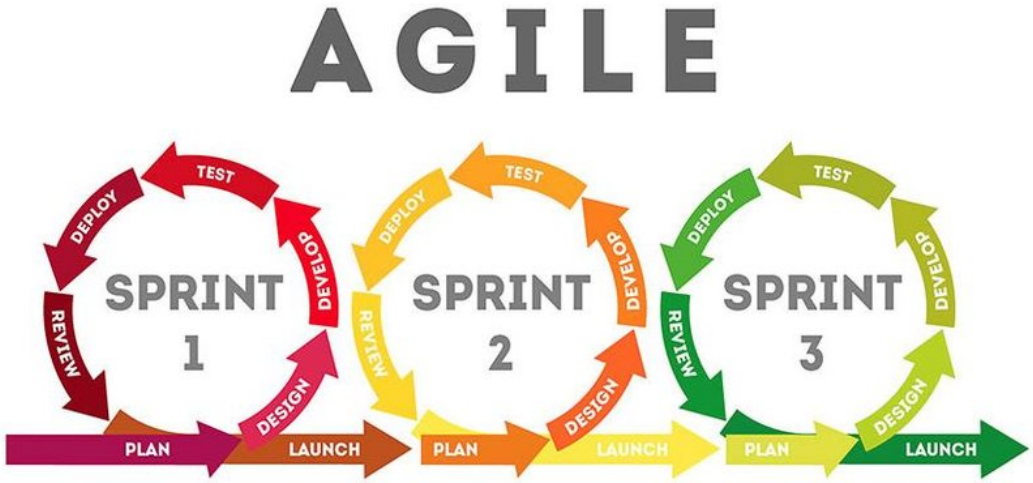


Figure 5.1: Agile Method
(Source: `https://www.researchgate.net/figure/Iterative-software-development-in-the-Agile-methodology-In-Table-1-some-of-the-fig2_353856176/`)

5.1.1.1 Scrum Framework

In adhering to the Scrum methodology, the project management approach adopted was incremental and iterative. This methodology involves dividing the project's lifecycle into smaller, manageable cycles known as sprints. These sprints are tackled independently, allowing for swift development and timely achievement of deadlines.

At the conclusion of each sprint cycle, the working software, referred to as an increment, is presented to stakeholders, typically colleagues and supervisors, for their feedback. This feedback loop ensures alignment with stakeholder expectations and facilitates course corrections as necessary.To maintain organization and clarity throughout the project, two essential backlogs are upheld. Firstly, the project backlog contains all tasks and to-dos pertaining to the entire project, serving as a comprehensive repository of requirements. Secondly, the sprint backlog is a subset of the project backlog, comprising only the tasks relevant to the ongoing sprint cycle. This focused backlog enables the team to concentrate effectively on the immediate objectives at hand.

5.1.2 Sprints

In this project, we have divided project into two sprints each of 8 weeks.

5.1.2.1 Sprint 1

During Sprint 1, our team focused on detailed research into the Transformer model and designed it to accommodate a single input. We concurrently developed a user interface featuring only text input and a simplistic design. Additionally, we constructed a TER model utilizing LSTM architecture, achieving an impressive accuracy rate of above 90%. Subsequently, we seamlessly integrated these components into a unified system.

Following completion of Sprint 1, we presented our progress to the stakeholder, our supervisor, and scheduled a meeting to obtain feedback. Upon review, it became apparent that our initial approach lacked efficiency. Consequently, we conducted a thorough examination of the outcome, identifying both necessary and unnecessary components. Armed with this insight, we commenced the next sprint with a refined focus, ready to address the shortcomings and optimize the system's performance accordingly.

5.1.2.2 Sprint 2

At the onset of the sprint, we conducted a comprehensive review of the system's deficiencies. To enhance efficiency, we collectively decided to incorporate a voice input feature. Recognizing the limitations of the Transformer model with only one input, we opted to introduce a new input: emotion. Additionally, we made the strategic decision to revamp the TER model for more effective emotion extraction. After experimenting with four different classifiers, we identified the most suitable one for integration.

Furthermore, we augmented the user interface by introducing new animations and facial features to enhance user experience. Subsequently, these enhancements were seamlessly integrated into the system. Following integration, we convened with the stakeholder to solicit feedback. As this marked the final sprint, we meticulously fine-tuned both the UI and backend to ensure optimal performance and user satisfaction.

We dedicated a total of 32 hours per sprint, as illustrated in the table below:

Table 5.1: Sprint Configuration

Sprint Length	**8 weeks**
Story point weight	1 story point = 1 hr
No.of working days per week	max 4 days
No.of working hours per day	2 hrs
No.of working days per sprint	max 24 days
No.of working hours per sprint	8*4 = 32 hrs

5.2 Project Management with Trello

5.2.1 Overview

Trello, a widely-used project management tool, played a pivotal role in organizing, tracking, and managing various aspects of our project. It provided a visual and collaborative platform that enhanced team coordination and streamlined project workflows.

5.2.2 Project Boards

We employed Trello's boards to structure our project into different phases, allowing for better organization and clarity. The following boards were utilized:

- **Backlog:** This board included all tasks, features, and ideas awaiting implementation.
- **In Progress:** Tasks actively being worked on were moved to this board, providing a real-time overview of ongoing activities.
- **Testing and QA:** Tasks ready for testing and quality assurance were managed on this board.
- **Completed:** Successfully implemented and tested tasks were moved to this board for archival purposes.
- **Ideas/Future Features:** This board captured additional ideas for potential future enhancements.

5.2.3 Detailed Lists for Each Phase

Each board included detailed lists for every phase, ensuring a granular breakdown of tasks and activities. For instance:

- **Backlog:** Tasks were categorized using labels (e.g., UI/UX, Development, Testing), and due dates were assigned where applicable.
- **In Progress:** Cards were moved here when active work commenced, team members were assigned, and due dates were monitored.

- **Testing and QA:** Cards moved to this list after development underwent thorough testing, and checklists within cards were employed for testing criteria.
- **Completed:** Successfully completed cards were archived to maintain a clean board.
- **Ideas/Future Features:** This list facilitated discussions on potential future features, encouraging collaboration.

5.2.4 Communication and Documentation

Trello served as a hub for team communication and documentation. Key boards included:

- **General Board:** A centralized location for project information, links, and resources, with comments facilitating task discussions and updates.
- **Documentation:** Dedicated to project documentation, including design specifications, user guides, and technical documentation.

5.2.5 Team Collaboration

Trello's collaborative features, such as comments, mentions, and attachments, played a crucial role in keeping team members informed and engaged. The calendar power-up was employed to visualize due dates and milestones.

5.2.6 Review and Reflection

Regular reviews and updates were conducted during team meetings, fostering continuous improvement and effective project management.

In summary, Trello proved to be an invaluable tool, offering a flexible and visual approach to project management that significantly contributed to the success of our project.

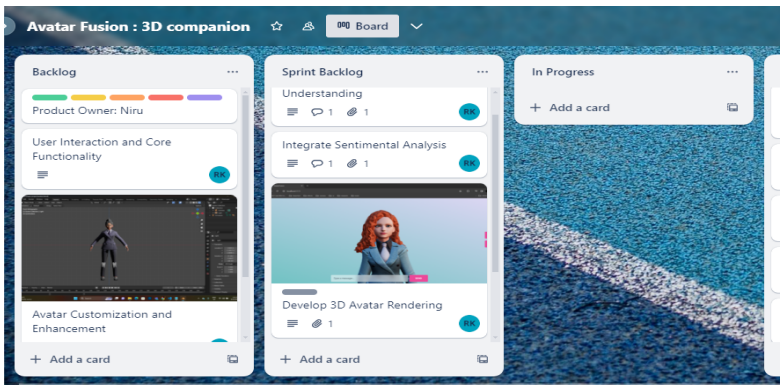


Figure 5.2: Trello as a Project Management Tool

Chapter 6

System Design and Architecture

6.1 Use Case Diagram

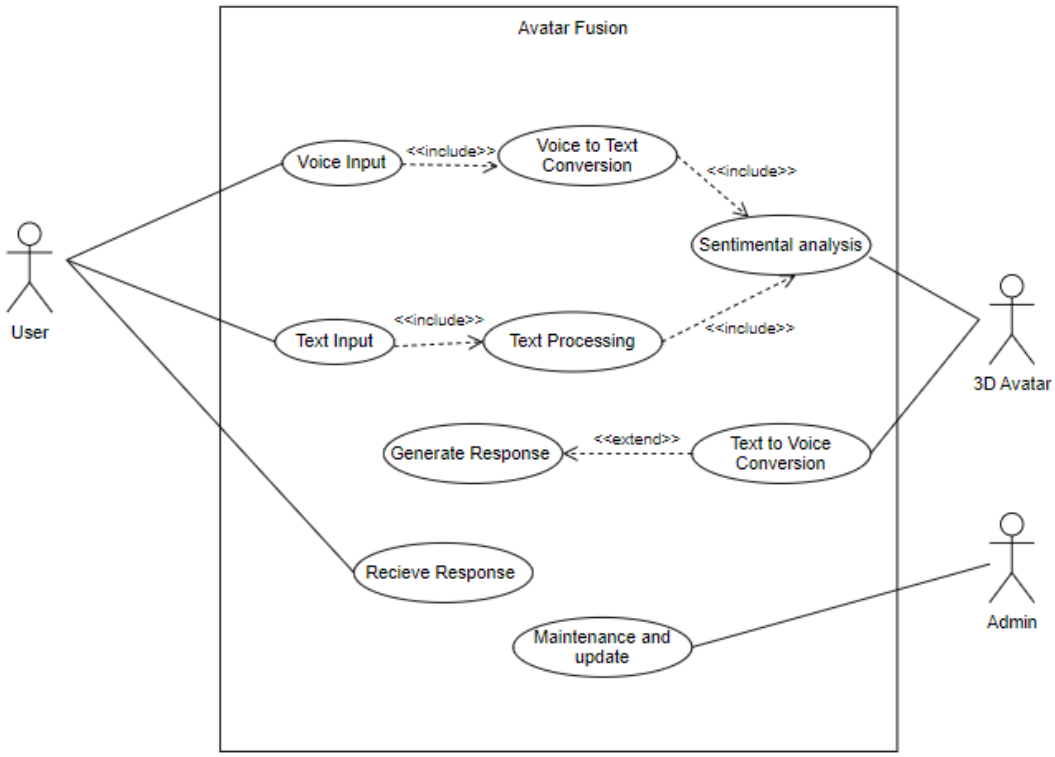


Figure 6.1: Use Case Diagram

In the Use Case Diagram, users input via voice or text.The voice to text conversion is followed accordingly. The inputted text and voice that is converted to the text will go under both natural language processing as well as sentimental analysis. Then the 3D avatar deliver the response to the input or the query provided by user. The administrator is responsible for maintenance and updating tasks.

6.2 Flowchart Diagram

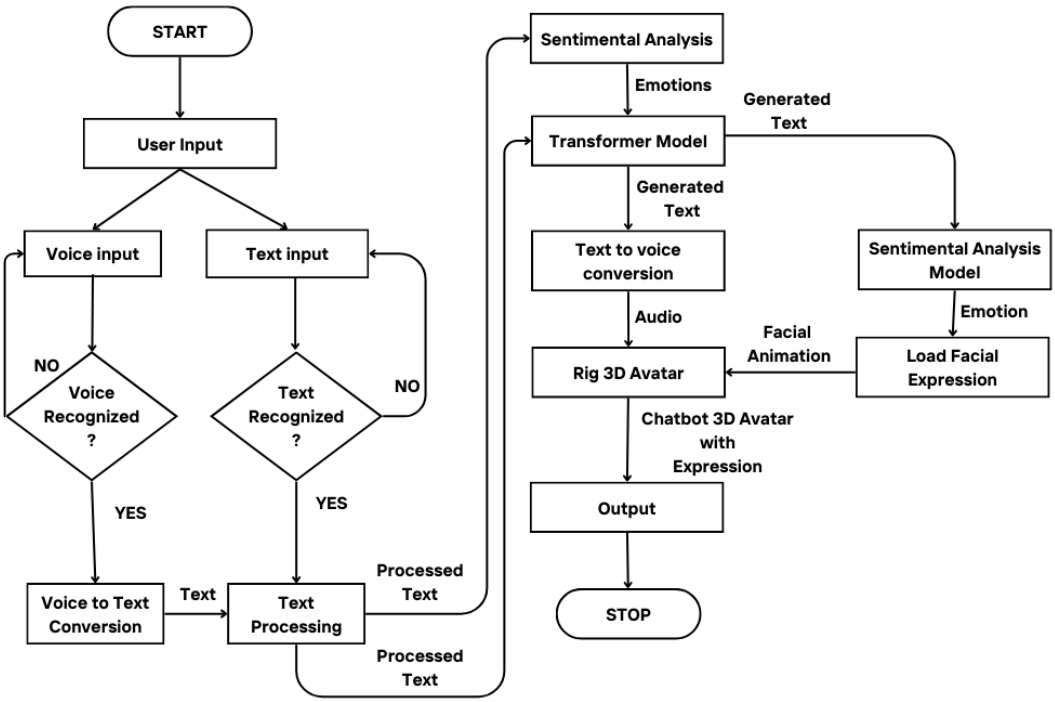


Figure 6.2: Flowchart Diagram

The user inputs data in the form of text or voice, with both undergoing recognition verification. In case of recognition failure, the user is prompted for input again. Text input proceeds directly to text processing. Voice input is converted to text, then undergoes text processing. The processed text will get operated in transformer model. The system generates voice output based on sentiments, rendering emotions in the avatar. If additional input is provided, the process iterates accordingly, ensuring a continuous interaction for an enhanced user experience.

6.3 System Block Diagram

The block Diagram of the proposed system is shown in the figure below:

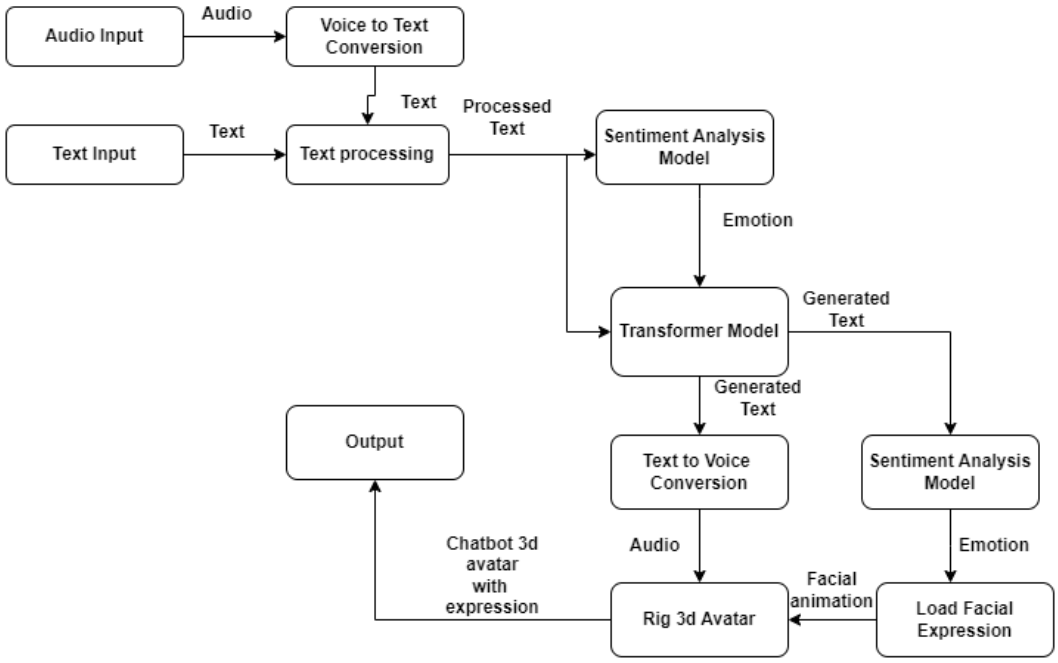


Figure 6.3: Block Diagram of AvatarFusion

Within this unified system, users engage through text and voice inputs. A proficient speech recognition module seamlessly translates voice inputs into text. The processed text, originating from both voice and text inputs, undergoes meticulous analysis before entering a transformer model. The resulting output from the transformer model serves pivotal roles in two modules: text-to-voice conversion and sentimental analysis. The former converts the generated text into audio, while the latter identifies emotional nuances. These components, coupled with facial emotion analysis, play a crucial role in the creation of a 3D avatar, providing users with a holistic and immersive experience that combines synthesized audio, emotional understanding, and dynamic visual representation.

6.4 Sequence Diagram

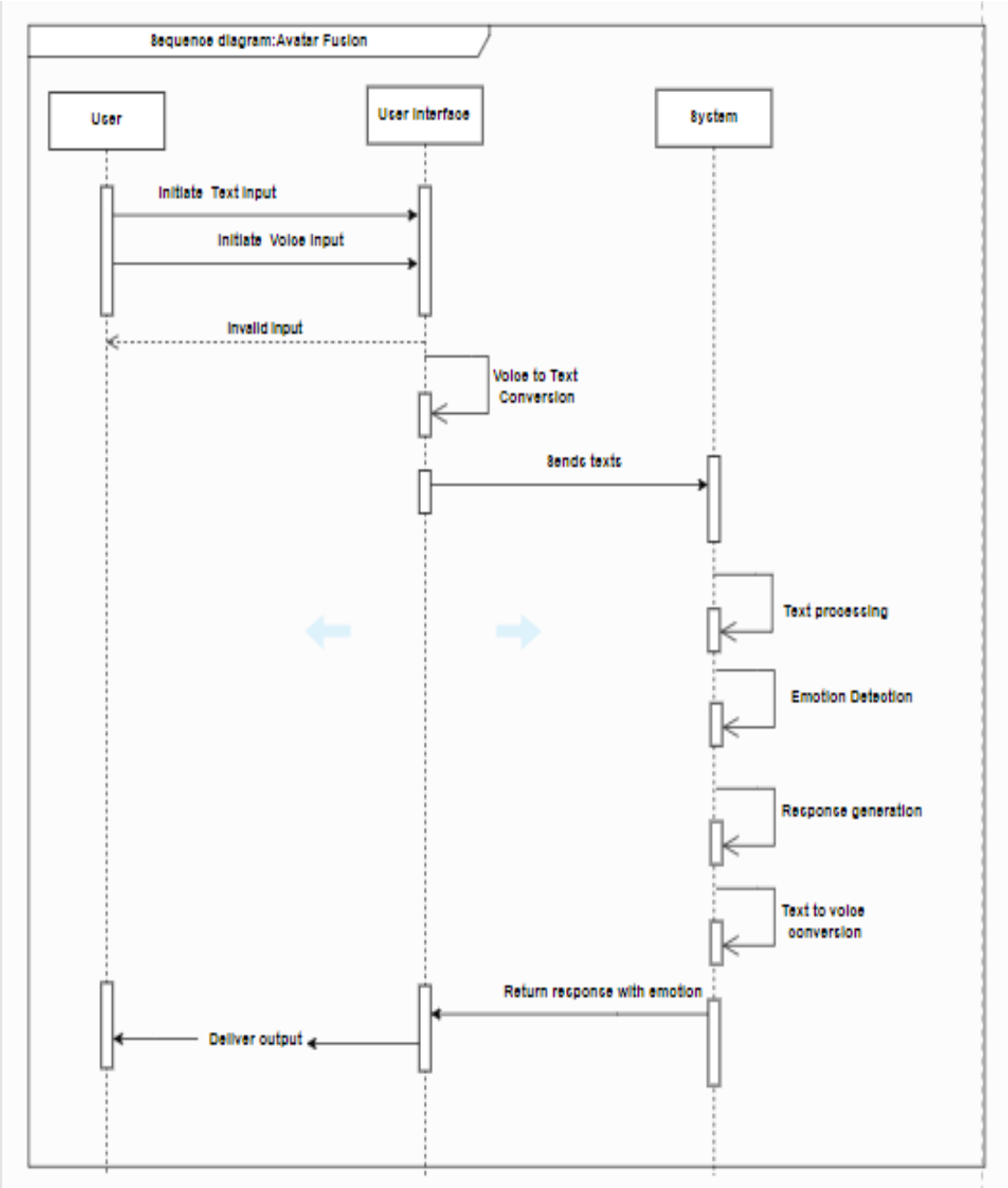


Figure 6.4: Sequence Diagram

Within this Sequence Diagram, users interact using both text and voice commands. A speech recognition system seamlessly translates spoken commands into text. The processed text, sourced from both voice and text inputs from User interface, undergoes thorough analysis before being fed into a transformer model in System. The resultant output from the transformer model fulfills crucial functions in two areas: converting text to speech and analyzing sentiment. The former translates the produced text into spoken words. These elements, along with facial emotion analysis, are integral to generating a 3D avatar, offering synthesized speech is given back to the users.

6.5 Model Description

To achieve a natural interaction wherein a chatbot dynamically responds to users' emotional cues, it is imperative to automatically adjust the emotion category based on the emotional context of the user's message. The proposed model aims to generate responses tailored to various emotions, including sadness, disgust, anger, happiness, neutrality, and fear, concurrently. Emotion detection can be conducted through text analysis, speech recognition, or facial expressions. In this project, we implemented emotion detection from text inputs only.

6.5.1 Text Emotion recognition Model

Text-based emotion detection can also be accomplished directly through textual analysis. This can be achieved through various methods, one of which is the rule-based approach. In this approach, specific grammatical and logical rules are established to identify emotions within documents. While creating rules for a few documents may be straightforward, dealing with large volumes of documents can introduce complexities.

Another approach is the employment of machine learning (ML) techniques to solve the emotion detection (ED) problem. Through the ML approach, texts are classified into different emotion categories using ML algorithms. This detection process commonly utilizes either supervised or unsupervised ML techniques. In our project, we opted for the ML method, where various machine learning algorithms were trained as multiclass classifiers and subsequently evaluated to determine the most effective one classifiers that we used in out project are:

6.5.1.1 Support Vector Machine

Support Vector Machines (SVMs) are widely utilized in text emotion detection models. They excel in classifying text into specific emotions like happy, sad, anger, or neutral. SVMs operate by identifying the hyperplane within the feature space that optimally segregates data points belonging to distinct classes.

To employ SVMs for text emotion detection, the initial step involves representing text as features. Commonly used feature representations include word embeddings and bag-of-words (BOW). Once the text has been converted into a feature representation, the SVM algorithm is trained on a labeled dataset to learn the mapping from features to emotions.

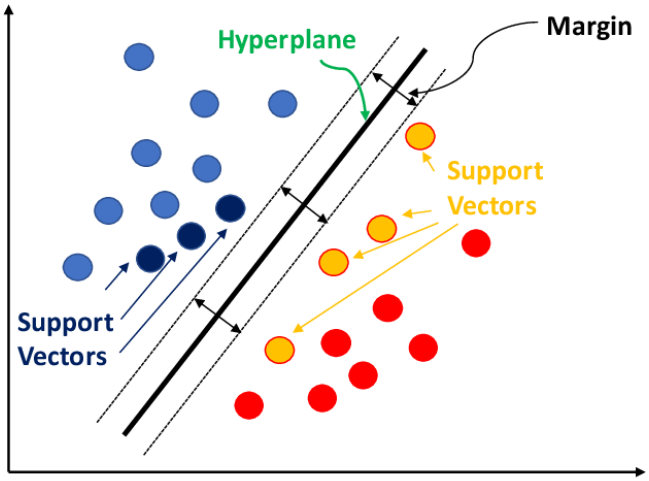


Figure 6.5: Diagram of Support Vector Machine
Source: Understanding Support Vector Machine [13]

6.5.1.2 Random Forest

The Random Forest algorithm is an excellent choice for text emotion detection projects due to its capability to manage large and intricate datasets, which are common in text analysis

.

When predicting the emotion of a new text, the algorithm utilizes the features of the text as inputs for each decision tree in the forest. It then records the predicted emotion for each tree. By aggregating the predictions from all the trees, the algorithm generates a final prediction. This ensemble approach helps mitigate the impact of noise and outliers in the data, leading to more accurate and robust predictions.

Another advantage of Random Forest is its ability to handle unbalanced datasets, a common occurrence in text emotion detection tasks where some emotions may have fewer samples than others. By balancing the samples during training, Random Forest can effectively address this issue. Additionally, Random Forest is adept at handling missing values in the data, which is advantageous when not all features are available for every text segment.

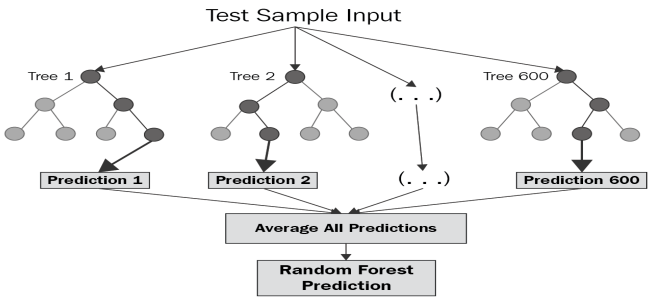


Figure 6.6: Diagram of Decision Tree
Source: Understanding Random Forest [14]

6.5.1.3 Multinomial NaiveBayes

The Multinomial Naive Bayes algorithm is commonly employed for text classification tasks due to its simplicity, speed, and capability to handle large datasets with numerous features. In the context of text emotion detection, this algorithm is utilized by first preprocessing the text data, typically by converting it into a bag-of-words (BOW) representation. Subsequently, the algorithm calculates the probability of each word in the vocabulary occurring within each emotion category. These probabilities are estimated using the training data.

To predict the emotion of new text data, the algorithm computes the probability of the text belonging to each emotion category based on the probabilities of the words present in the text. It then selects the emotion category with the highest probability as the predicted emotion for the given text.

6.5.1.4 Voting Classifier

The concept behind a Voting Classifier involves aggregating the outputs of multiple models to make a consolidated prediction. This method combines the predictions of individual classifiers passed into the Voting Classifier, ultimately determining the output class based on the majority of votes. Instead of creating separate models and evaluating their accuracies independently, a single model is trained using these diverse models, predicting outputs based on their collective majority voting for each output class. There exist two primary types of Voting Classifier: hard voting and soft voting.

In hard voting, the final prediction is determined by the majority vote among the predictions of the constituent models. Conversely, in soft voting, the final prediction is based on the average of the probabilities assigned by the individual models.

The text emotion prediction is finalized using the voting classifier algorithm, wherein the outputs of Random Forest, Support Vector Machine, and Multinomial Naive Bayes algorithms are combined to produce the ultimate emotion detection outcome in our model.

6.6 Generative Model

The generative model utilized by the chatbot doesn't rely on pre-existing repositories. Instead, it leverages advanced deep learning techniques to generate responses to queries. This approach parallels machine translation, where source code is translated from one language to another, except here, the input is transformed into an output. Through machine learning models, the chatbot's functionality is greatly enhanced, enabling it to recognize numerous questions posed by humans, fostering more insightful and dynamic interactions.

The advent of transformer models in the realm of neural machine translation has been notable, with successful adaptations to dialogue-related challenges. Specifi-

cally, a transformer model is employed to construct the generative model, facilitating the chatbot's ability to generate responses in a conversational context.

6.6.1 Transformer Model

Before the advent of transformers, recurrent neural networks (RNNs) were commonly used in natural language processing tasks. RNNs process sequences of words by sequentially handling each word and incorporating the result into the processing of the next word. This allows RNNs to maintain context throughout a sentence, rather than treating each word in isolation. However, RNNs had limitations that hindered their effectiveness. Firstly, they were slow because they processed data sequentially, unable to leverage parallel computing hardware like graphics processing units (GPUs) for training and inference. Secondly, RNNs struggled with processing long sequences of text, as the impact of the initial words in a sentence diminished as the network progressed through the text.

The introduction of transformer models addressed these drawbacks. A transformer is a neural network architecture that learns context and meaning by capturing relationships within sequential data, such as the words in a sentence. Proposed in the paper "Attention is All You Need," [15] transformers rely solely on self-attention mechanisms and are highly parallelizable. Transformers revolutionized sequential deep learning models in two significant ways. Firstly, they enabled the parallel processing of entire sequences, dramatically increasing the speed and capacity of sequential models. Secondly, transformers introduced attention mechanisms, allowing the model to capture relationships between words across long text sequences in both forward and reverse directions.

Transformers revolves around three key concepts:

1. Positional Encodings: Since Transformers lack recurrence or convolution, positional encoding is introduced to provide the model with information about the relative positions of words within a sentence. The positional encoding vector is added to the embedding vector. While embeddings represent tokens in a d-dimensional space where tokens with similar meanings are closer, they do not encode the relative positions of words in a sentence. Therefore, after incorporating positional encoding, words in the d-dimensional space become closer to each other based on both their meaning and position in the sentence.
2. Attention: Attention allows the model to focus on specific parts of the input data while disregarding others, aiding in solving the task effectively. In tasks like machine translation, humans naturally focus more on certain parts of a sentence, such as who, when, and where.
3. Self-Attention: Self-attention, also known as intra-attention, facilitates the creation of connections within the same sentence. This mechanism helps neural networks disambiguate words, perform tasks like part-of-speech tagging and entity resolution, and learn semantic roles.

The Transformer architecture consists of two main components: the Encoder and the Decoder.

1. Encoder: This module compresses an input string from the source language into a vector representation, capturing the relationships between words.
2. Decoder: The decoder module transforms the encoded vector into a string of text in the destination language, facilitating the translation process.

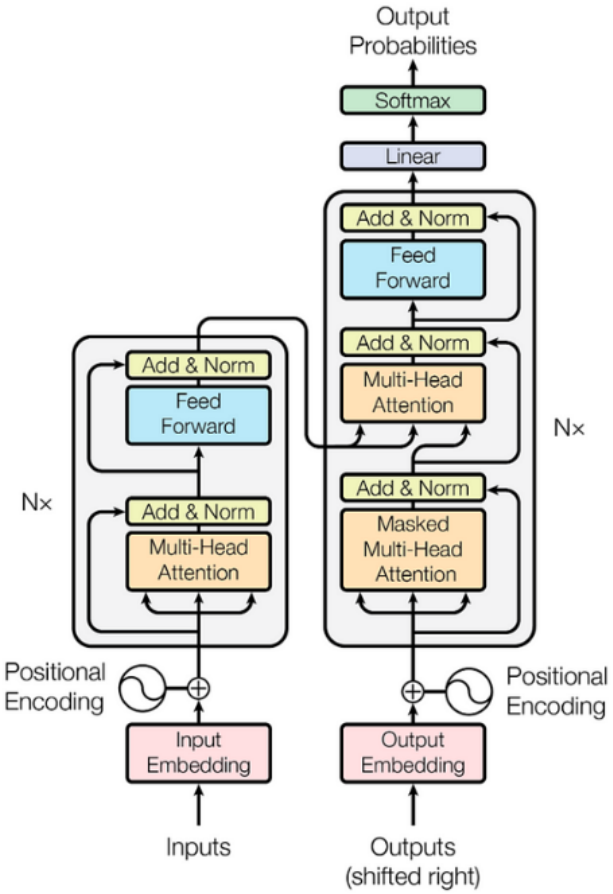


Figure 6.7: Illustration of a Transformer Model
Source: Attention is all you need [16]

6.7 Work Flow Description

The project basically is composed of two intelligent models.

1. Text Emotion Recognition Model
2. Generative Model

6.8 Dataset Accumulation and Preprocessing

The dataset plays a pivotal role in the development of a conversational bot. The chatbot's responses are predominantly influenced by the dataset it learns from during training. Currently, there are datasets specifically tailored for developing emotionally aware chatbots. These datasets are accessible in English and other languages. They are compiled from diverse sources such as social media platforms, online websites, and manually constructed through crowdsourcing efforts. These datasets serve as valuable resources for training chatbots to recognize and appropriately respond to users' emotions during interactions.

The dataset majorly contain empathetic conversations along with emotions of user's response. Since this is an NLP project, most of the datasets were obtained from huggingface.com

6.8.1 Text Emotion Recognition

6.8.1.1 Data Collection

For our Text emotion recognition model we collected different dataset from different sources like kaggle,github repository toronto etc.

One of the datasets used is the Tweet Emotion dataset, sourced from Kaggle [17]. This dataset comprises English Twitter messages annotated with six basic emotions: anger, fear, joy, love, sadness, and surprise. The authors of the dataset also curated a separate collection of English tweets from the Twitter API, categorized into eight basic emotions, including anger, anticipation, disgust, fear, joy, sadness, surprise, and trust.The dataset consists of three files: train, test, and valid. These files were merged and further preprocessed. While the dataset originally contained eight basic emotions, only the data pertaining to the required six emotions were selected for the project.

Our second dataset was sourced from Hugging Face [18]. This dataset consisted of 34,791 entries, encompassing eight emotions: joy, sadness, fear, anger, surprise, neutral, disgust, and shame. From this dataset, we extracted the required data pertaining to six emotions.

The third dataset comprises 40,000 records featuring 13 distinct emotions. This dataset, sourced from the public domain platform "Sentiment Analysis in Text" on data.world, consists of tweets annotated with their corresponding emotions [19]. Each record includes three columns: tweet ID, sentiment, and content. The "content" column contains the original tweet text, while the "sentiment" column specifies the emotion conveyed in the tweet. From the original set of 13 emotions, we narrowed down the data selection to focus on six specific emotions.

Furthermore, our final dataset, obtained from Hugging Face and known as the GoEmotions dataset, comprised 58,000 meticulously curated Reddit comments. These comments were labeled for 27 emotion categories, including Neutral, as

detailed in the corresponding paper [20]. Due to limited availability after merging, we only retained data pertaining to the "disgust" emotion category from this dataset.

6.8.1.2 Data Merging

After extracting all the required datasets, we narrowed our focus to only six emotions for our basic emotion classification model, namely: sad, anger, fear, happy, neutral, and disgust. Despite encountering a total of 16 emotions across all datasets, we specifically chose these six for our classification model, as illustrated in the figure below.

```
sadness       17684
neutral       10892
worry          8459
fear           7783
anger          7116
joy            6761
disgust        6157
love           5483
happiness      5209
surprise       2906
fun            1776
relief         1526
hate           1323
empty           827
enthusiasm      759
boredom         179
Name: Emotion, dtype: int64
```

Figure 6.8: Emotion Detected in Dataset

Subsequently, we amalgamated these emotions into six basic categories by grouping them according to their similarities.

- sadness was renamed to emotion "sad"
- happiness,love,joy were grouped under the emotion "happy"
- worry was renamed to the emotion "fear"
- disgust was kept as the emotion "disgust"
- anger,hate were grouped under the emotion "angry"
- enthusiasm,relief were grouped under the emotion "neutral"

Hence, our consolidated dataset can be seen as follows:

SN	Folder	Number of data
1	Sad	17684
2	happy	17453
3	fear	16242
4	neutral	13177
5	angry	8439
6	disgust	6157

Table 6.1: Text Emotion Classes Information after Merging

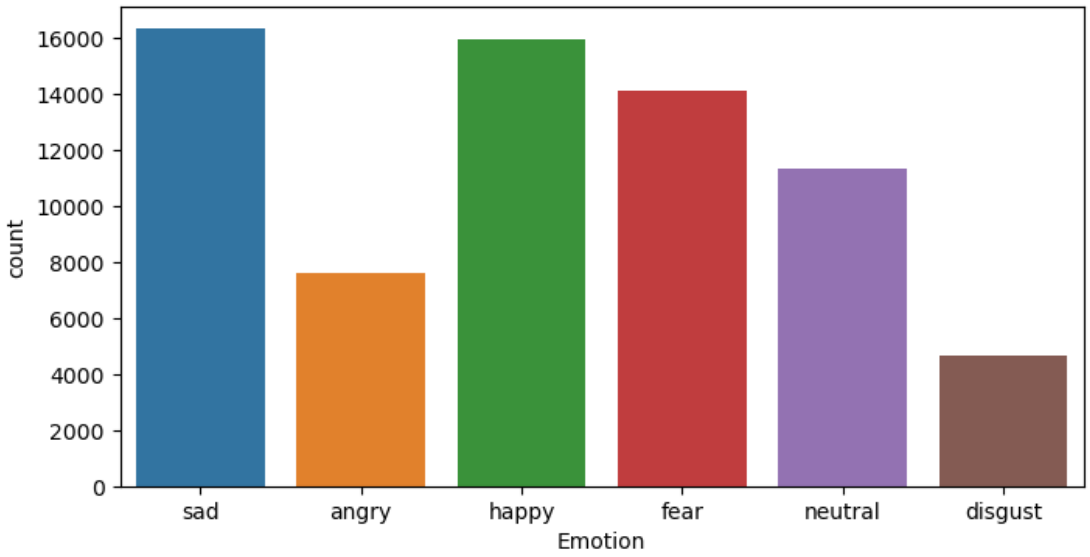


Figure 6.9: Merged Data distribution

6.8.1.3 Data cleaning

For cleaning the data we followed following steps:

1. Remove stopwords: The stopwords such as "the", "a", "an", "and", "of" and others were removed
2. Remove special characters
3. Lowercase: All the texts were lowercased.
4. Handle Duplicate data: More than 2680 duplicate data were obtained which were completely removed.

6.8.1.4 Data Augmentation

Number of datasets were imbalanced so we augmented the data using nplaug library for three emotions, angry, digust and neutral.The "nlpaug" module was imported, enabling augmentation based on text synonyms. These synonyms are generated from WordNet, which serves as a comprehensive lexical database of English. Finally, total data of 87092 was obtained. Thus our data augmented data can be visualized as:

SN	Folder	Number of data
1	Sad	16186
2	happy	15854
3	fear	13996
4	neutral	13699
5	angry	14359
6	disgust	12998

Table 6.2: Emotion Classes Information after Augmentation

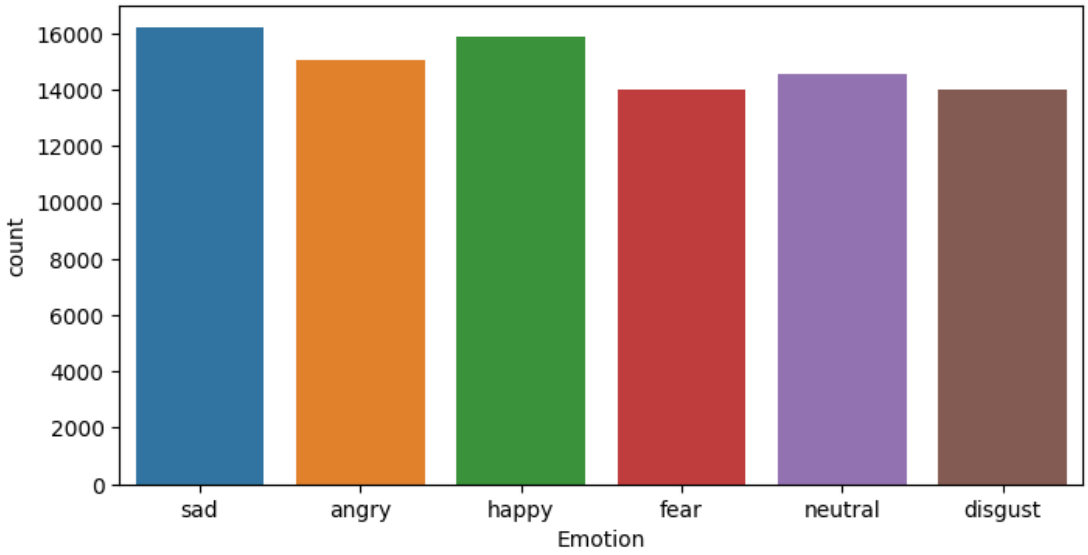


Figure 6.10: Augmented Data distribution

6.8.1.5 Vectorization

For the TER model, we explored the utilization of TF-IDF vectorization as a feature extraction technique during the model training phase. TF-IDF, short for term frequency-inverse document frequency, amalgamates the concept of term frequency (how often a term appears within a document) and inverse document frequency (how unique a term is across multiple documents). This approach enables the identification of both the most common terms within a document and the rare terms across the entire corpus. As a result, TF-IDF facilitates the selection of distinctive term vectors as feature sets for model training.

TF-IDF Vectorizer functions by transforming text inputs into feature vectors, enabling their use as input for model estimators. It constructs a vocabulary or dictionary that maps each token (word) to a feature index based on its frequency in the matrix, with each unique token assigned a feature index. TF-IDF was chosen for its capacity to represent text data as a collection of vectors, which can be readily employed as input for various machine learning models.

6.8.2 Generative model

6.8.2.1 Data Collection

We sourced the dataset for our generative model from the Hugging Face website [21], originally derived from the research paper titled "Towards Empathetic Open-domain Conversation Models: a New Benchmark and Dataset" [22]. This dataset consists of 25,000 conversations rooted in emotional contexts and is publicly accessible for empathetic dialogues. The data collection involved a two-step process: the first worker presented a prompt, and the second worker responded with a message demonstrating empathy towards the prompt. This dataset stands out for its emphasis on empathy, a crucial aspect of authentic and impactful human interaction.The sample of data can be seen here A.3.It consist of 8 labels: conv id, utterance idx, context, prompt, speaker idx, utterance, selfeval, tags.

Furthermore, we acquired a dataset from Kaggle [23], which includes around 69,000 instances of empathetic dialogue. This dataset is structured with four labels: situation, emotion, dialogue, and labels. The "labels" attribute specifically contains responses to the dialogues as shown in A.3.

6.8.2.2 Merging and Classification

Merging the dialogues, replies, and emotions from second datasets with context, utterance, conv_id of first dataset, we compiled approximately 200,000 entries in our CSV file, as shown in A.4. Since our project focuses solely on six basic emotions, we converted all the emotions into these basic categories based on their similarities:

- Sentimental, sad, nostalgic, lonely, devastated, disappointed were grouped under the emotion "sad"
- Proud, impressed, joyful, surprised, excited, grateful, happy were grouped under the emotion "happy"
- Apprehensive, anxious, terrified, afraid were grouped under the emotion "fear"
- Ashamed, guilty, embarrassed, disgusted were grouped under the emotion "disgust"
- Angry, annoyed, jealous, furious were grouped under the emotion "angry"
- Faithful, trusting, caring, hopeful, anticipating, prepared, confident, content were grouped under the emotion "neutral"

The table below illustrates the distribution of data across various emotions.

The table below displays the total dataset sizes for both the emotion recognition model and the generative model.

SN	Emotion	Number of data
1	Neutral	31775
2	Happy	34966
3	Sad	34181
4	Angry	32821
5	Fear	30835
6	Disgust	31682

Table 6.3: Generative Data Emotion Feature Information

S.N.	Dataset for	Data	Number of Data
1	TER model	text	87908
2	Transformer Model	text	206260

Table 6.4: Description of Custom Dataset

6.8.2.3 Feature selection

The dataset underwent extraction of only the necessary features/columns, discarding irrelevant ones such as "selfeval" and "tags" etc. The retained features are as follows:

1. context
2. utterance
3. conv_id

these Features were copied into new csv file and renamed as:

1. context was kept as it is context.
2. utterance was renamed to response.
3. conv_id was renamed to emotion.

6.8.2.4 Data Cleaning

Similarly, for the Generative Model, we performed data cleaning on the merged dataset for the text emotion dataset. This involved lemmatizing, lowercasing, and removing noise data to ensure the dataset's quality and consistency.

1. Lemmetization: Lemmatization goes beyond simple word reduction by considering a language's entire vocabulary and applying morphological analysis to words. Its aim is to remove only inflectional endings, returning the base or dictionary form of a word, known as the lemma. In our dataset sample see A.4, the "context" column underwent lemmatization using WordNetLemmatizer with POS tagging. POS tagging involves marking each word in a text with its corresponding grammatical category, such as noun, verb, adjective, etc. This information is then utilized to determine the lemma of each word. For instance, the words "run", "running", and "ran" are all transformed into "run".

2. Handling nan values: None of the columns had empty values except for "history" column. The cells with empty value in histories marked starting of conversation so, it was filled with tag SOC .
3. Lowercasing: All the words were converted into lowercase because the two words that mean the same but when not converted to the lower case. Those two are represented as two different words in the vector space model (resulting in more dimensions).
4. Remove noisy data: As there were some noisy data in "emotion" column which could not be modified and made of use, so those rows were removed. We were able to obtained total of 26393 number of clean data. The data contained following number of conversations in each emotion.
5. Removing Special Characters: The text data were then preprocessed to remSove unique symbols such as '?', '.', '!', '-', '_', "̂ and other unwanted symbol and spaces

6.8.2.5 Tokenization

After cleaning the datasets as the emotion model, tokenization was used. Tokenization involves dividing text into tokens, which are subsequently converted into numerical representations for use by machine learning models during processing and training. In our project, subword tokenization was implemented. This method preserves frequently occurring words without further division, while breaking down rare words into meaningful subwords. As a result of implementing subword tokenization, we achieved a total vocabulary size of 8262, inclusive of start and end tokens.

6.8.3 Model Creation

6.8.3.1 Sentiment Analysis Model

In the project, we developed a text emotion recognition model utilizing various machine learning algorithms, including Random Forest, SVM, MultinomialNB, and a Voting classifier. For each algorithm, specific parameter values were set, while other parameters were left at their default settings:

1. Random Forest: 'random_state = 0'
2. SVM: 'kernel='sigmoid'', 'random_state = 0', 'probability=True'
3. MultinomialNB: default parameters

All algorithms were imported from the 'scikit-learn' module in Python. The dataset was divided using an 80-20 split, with 80% of the data used for model training.

6.8.3.2 Evaluation metrices for TER Model:

1. **Confusion Matrix:** A Confusion matrix is a pivotal evaluation tool in classification problems, encapsulating a model's performance by detailing true positives, true negatives, false positives, and false negatives. It serves as

the basis for calculating key metrics like accuracy, precision, recall, and F1-score, offering insights into the model's efficacy and areas for improvement. This matrix is indispensable for refining algorithms, identifying misclassifications, and enhancing the overall predictive capabilities of classification model.

2. **Accuracy:** The 'accuracy' metric is a commonly used evaluation metric in classification problems. It measures the proportion of correctly classified samples among all the samples in the test set. Here is an overview of the algorithm used to calculate accuracy:

   Step 1: Feed the test data through the trained model to get the predicted class labels.

   Step 2: Compare the predicted class labels with the true class labels.

   Step 3: Calculate the proportion of correctly classified samples as the accuracy.

   Step 4: Repeat steps 1-3 for all the test samples.

   Step 5: Calculate the average accuracy over the test set.

   The formula for accuracy is:

$$\text{Accuracy} = \frac{\text{Number of correctly classified samples}}{\text{Total number of samples}}$$

3. **Precision:** Precision is a statistical metric that measures the proportion of true positive results among the total positive results in a classification or detection task. In other words, it is a measure of the accuracy of positive predictions or decisions made by a model or a system. Mathematically, it can be expressed as:

$$\text{Precision} = \frac{TP}{(TP+FP)}$$

   where:

   TP = True Positive (it is the number of cases where the model correctly identifies a sample as belonging to a specific class when it does indeed belong to that class.)

   FP = False Positive (it is a case where the model mistakenly identifies a sample as belonging to a specific class when it actually belongs to a different class.)

4. **Recall:** Recall, also known as sensitivity or true positive rate (TPR), is a statistical metric that measures the proportion of true positive instances that are correctly identified as positive by a classification or detection system. In other words, it is a measure of how well the system is able to identify all instances of the positive class. Mathematically, it can be expressed as:

$$\text{Recall} = \frac{TP}{(TP+FN)}$$

where:

TP = True Positive (it is the number of cases where the model correctly identifies a sample as belonging to a specific class when it does indeed belong to that class.)
FN = False Negative (it is a case where the model mistakenly identifies a sample as not belonging to a specific class when it actually does belong to that class.)

5. **F1 score:** The F1 score is a statistical metric used to evaluate the performance of a classification or detection system. It is a harmonic mean of precision and recall, and provides a single score that balances the trade-off between these two metrics. Mathematically, it can be expressed as:

$$\text{F1 Score} = \frac{2 * (precision * recall)}{(precision + recall)}$$

6.8.3.3 Generative Model

The Transformer was proposed in the paper "Attention is All You Need" [16]. This transformer chatbot architecture is based on article created by Harvard's NLP group called "The Annotated Transformer". For generative model, Transformer model with multihead attention was used. The work description about the transformer model is already discussed before. The dataset is preprocessed as discussed before with features selections and additions A.4. The architecture of the transformer was customized to incorporate all the five features/columns of our dataset. Also, three Multihead attentions were used. The custom transformer model is trained to learn the type of replies and vocabularies used in responses with their corresponding set of features and input messages.

A custom layer PositionalEncoding layers was also added. This layer implements the positional encoding used in the Transformer model. Various other layers such as encoding layer, decoding layers and transformer models were customized to fit our project.The model architecture of the encoder incorporates three input features i.e. messages, emotions and replies referenced as input1, input2, input3.

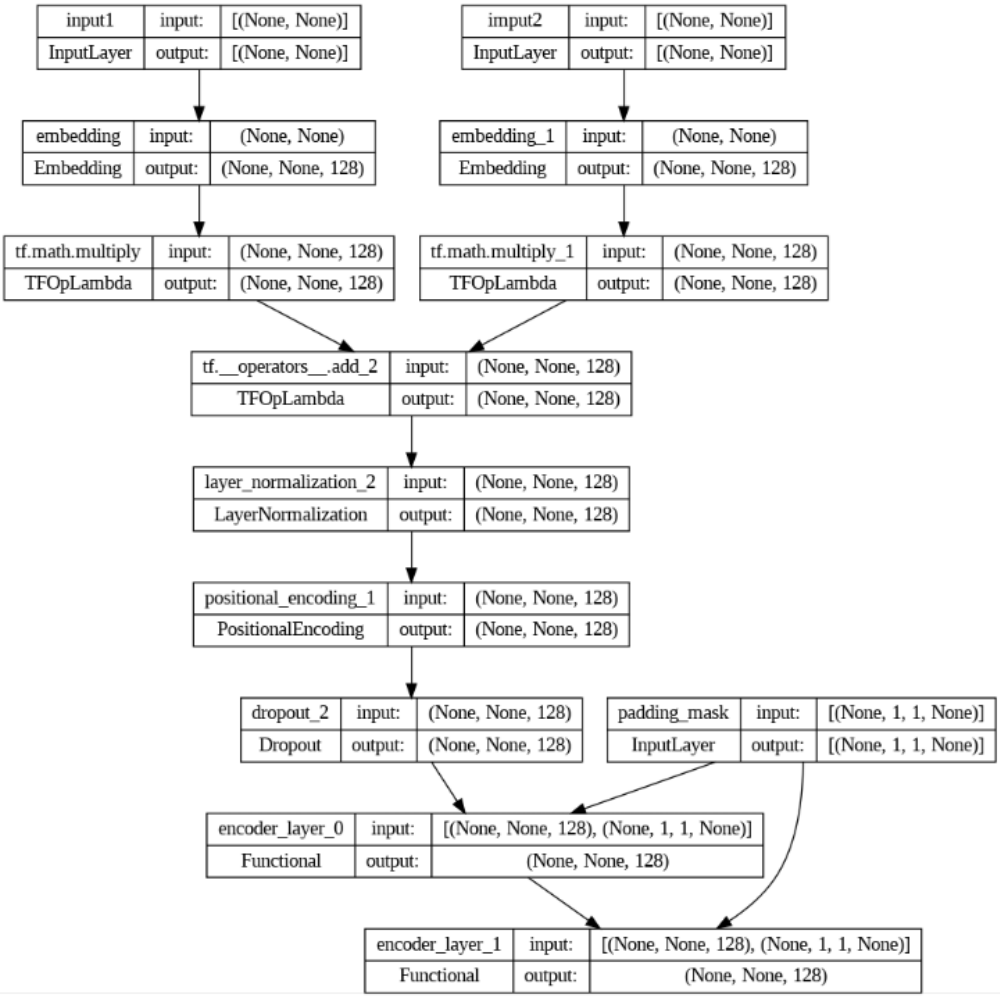


Figure 6.11: Encoder Architecture

Similarly, figure below are the visualization of the decoder architecture and custom transformer architecture for the project. Decoder includes layers such as look ahead mask, padding masks, decoder layers and others.

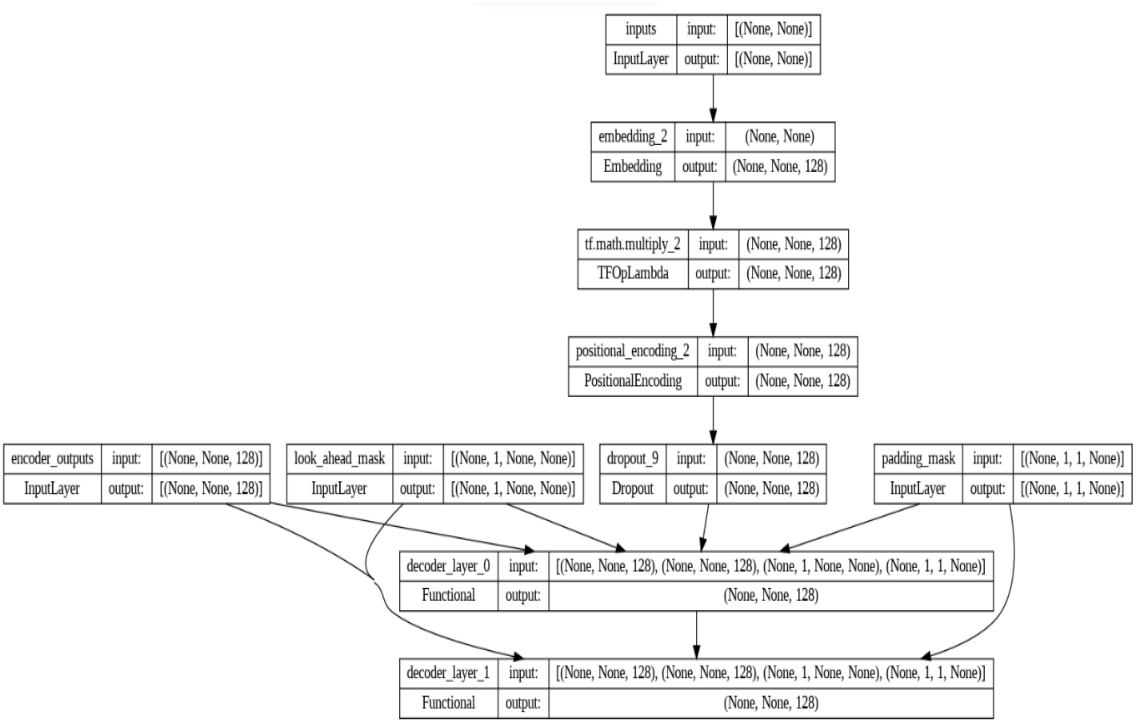


Figure 6.12: Decoder Architecture

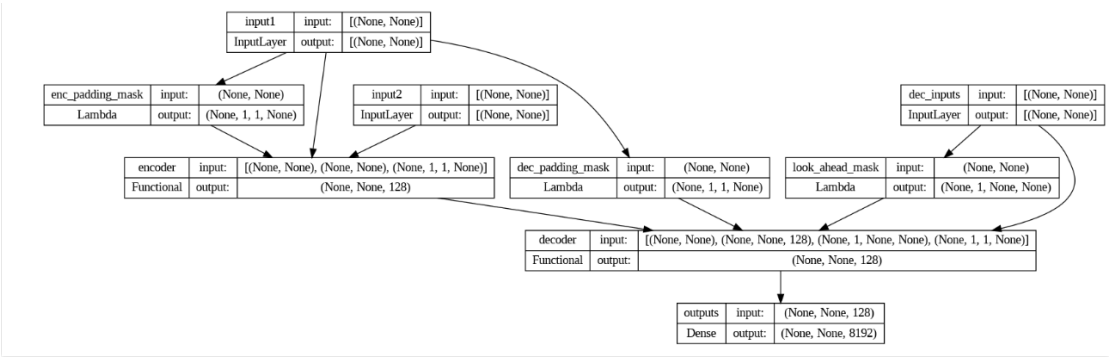


Figure 6.13: Transformer Architecture

6.8.3.4 Algorithms Used in Generative Model

- **Adam Optimizer:** Adam (short for Adaptive Moment Estimation) is a stochastic gradient descent optimization algorithm that combines ideas from both the Adagrad and RMSProp algorithms. The basic idea behind Adam is to compute adaptive learning rates for each parameter based on the first and second moments of the gradients.

  Step 1: Initialize the parameters of the model.

  Step 2: Initialize the first and second moments of the gradients to zero.

  Step 3: For each iteration:

  a. Compute the gradients of the loss with respect to the parameters.

b. Update the first moment of the gradients using a moving average.

c. Update the second moment of the gradients using a moving average.

d. Compute the adaptive learning rate for each parameter.

e. Update the parameters using the adaptive learning rates and the first and second moments of the gradients.

Step 4: Repeat until convergence or for a fixed number of iterations.

6.8.4 Evaluation metrices for Transformer model:

1. Perplexity: Perplexity is a measure commonly used in natural language processing and information theory to evaluate the performance of language models. It provides a quantitative measure of how well a probability distribution or a language model predicts a sample. It can be calculated as,

$$\text{Perplexity} = 2^{\frac{-1}{N}\sum_{i=1}^{N}\log_2 \text{P}(xi)}$$

where as,
N is the number of words or events in the dataset.
P(xi) represents the probability assigned by the model to the occurrence of the i-th word or event.

2. BLEU score: The BLEU (Bilingual Evaluation Understudy) score is a metric used to evaluate the quality of machine-generated translations by comparing them to one or more reference translations. The BLEU score ranges from 0 to 1, with a higher score indicating a better match to the reference translations.
It is calculated as,

$$\text{BLEU} = \text{BP} \times \exp\left(\frac{1}{n}\sum_{i=1}^{n}\log p_i\right)$$

where as,
BP is The Brevity Penalty (BP) is calculated as $\min\left(1, \frac{\text{output length}}{\text{reference length}}\right)$.
n represents the maximum n-gram order considered, typically set to 4.
The precision p_i for n-grams is calculated as the ratio of the number of n-grams in the candidate translation that match a reference translation to the total number of n-grams in the candidate translation.

6.9 3D Avatar Development

In our project, we've enhanced user interactions by incorporating a chatbot that facilitates engaging and friendly conversations. To bring a dynamic and personalized touch, we utilized Blender for 3D avatar modeling, creating animations for expressions such as smiles, sadness, and anger. These animations were then

exported as an FBX file. Leveraging the capabilities of React Three Fiber, we seamlessly integrated these 3D avatars into our user interface. The result is an interactive and visually appealing platform where users can enjoy conversations with expressive avatars, enhancing the overall user experience in our chatbot-driven environment.Screenshot of 3D avatar model is shown in A.6.

6.10 UI Development

The interface design utilized React Three Fiber for its capability to load 3D models and animations seamlessly. The 3D model, crafted in Blender, was successfully integrated into the React development environment. Leveraging the functionalities provided by Three.js libraries, such as useFrame and useEffect, animations were loaded and dynamically altered based on responses. This approach allowed for responsive and interactive animations within the interface. The resulting user interface, illustrated in the accompanying figure, reflects the successful fusion of React Three Fiber, Blender-generated 3D assets, and dynamic animations, showcasing a visually compelling and responsive environment.The UI developed can be seen in A.7

6.11 Model Integration

Once all models and the user interface (UI) were constructed and saved, they were seamlessly integrated using FastAPI. Initially, Whisper AI was considered for integration, but later, the WebKitSpeechRecognition module was discovered. This module facilitates the conversion of speech into text directly from the User Interface, which is then forwarded to the backend. Leveraging FastAPI, we unified the generative model and emotion recognition model. In this API setup, the user's message undergoes emotion prediction via the emotion recognition model. Subsequently, both the output of the emotion recognition model and the user's message are fed into the generative model. The generative model generates textual responses to the user, which are then analyzed by the emotion recognition model to ascertain the avatar's expression in the 3D space. Audio is generated utilizing the ElevenLabs API, followed by the generation of a lipsync file for the 3D avatar using the Rhubarb application. Finally, the emotion of the avatar, along with the audio and lipsync file, is transmitted to the UI for display.

Chapter 7

Experiments

7.1 Text Emotion Recognition

In our investigation of the TER model, we experimented with four different classifiers, carefully observing their performance and the outcomes they produced.

7.1.1 For Voting Classifier

1. **ROC curve of Voting Classifier:**

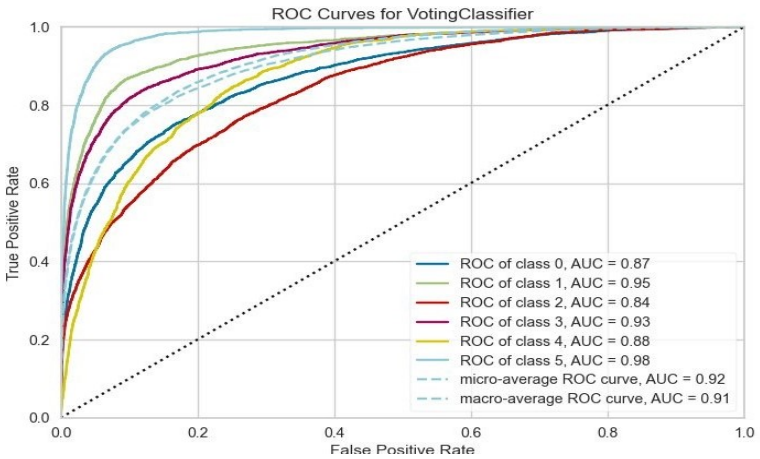


Figure 7.1: ROC-AUC curve of Voting Classifier

2. **Classification Report of Voting Classifier:**

```
Accuracy:  0.6648105625717566

Model MSE:  2.1072330654420206
Model MAE:  0.7412169919632606
              precision    recall  f1-score   support

           0       0.61      0.64      0.63      3207
           1       0.78      0.74      0.76      2835
           2       0.63      0.42      0.50      2913
           3       0.70      0.75      0.72      3116
           4       0.51      0.60      0.55      2746
           5       0.79      0.86      0.82      2603

    accuracy                           0.66     17420
   macro avg       0.67      0.67      0.66     17420
weighted avg       0.67      0.66      0.66     17420
```

Figure 7.2: Classification report of Voting Classifier

3. **Confusion Matrix of Voting Classifier:**

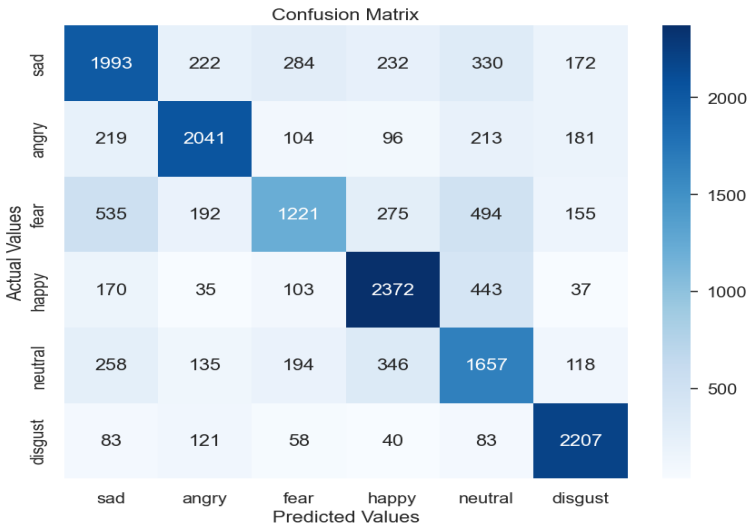


Figure 7.3: Confusion Matrix of Voting Classifier

7.1.2 For Naive Bayes Classifier

1. **ROC curve of Naive Bayes Classifier:**

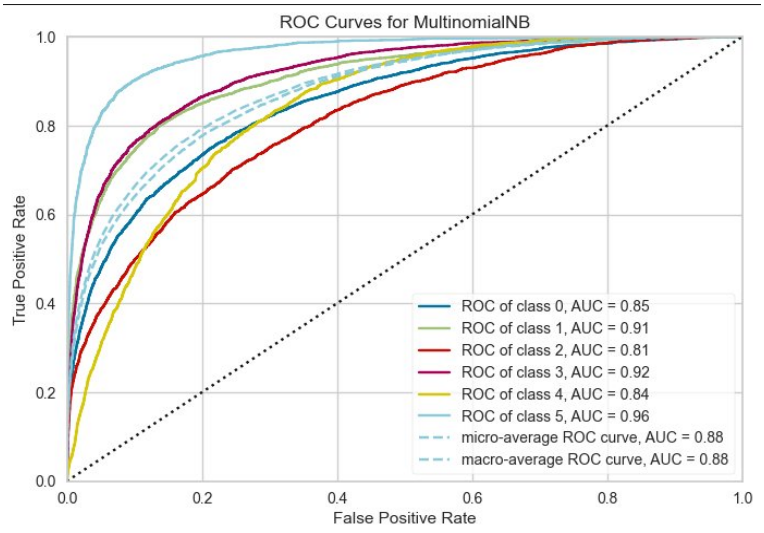


Figure 7.4: ROC-AUC curve of Naive Bayes Classifier

2. **Classification Report of Naive Bayes Algorithm:**

Accuracy: 0.5994259471871413

Model MSE: 2.6757749712973595
Model MAE: 0.9090700344431688

	precision	recall	f1-score	support
0	0.47	0.71	0.56	3207
1	0.67	0.68	0.68	2835
2	0.60	0.35	0.44	2913
3	0.60	0.76	0.67	3116
4	0.57	0.32	0.41	2746
5	0.78	0.76	0.77	2603
accuracy			0.60	17420
macro avg	0.62	0.60	0.59	17420
weighted avg	0.61	0.60	0.59	17420

Figure 7.5: Classification report of Naive Bayes Algorithm

3. **Confusion Matrix of Naive Bayes Classifier:**

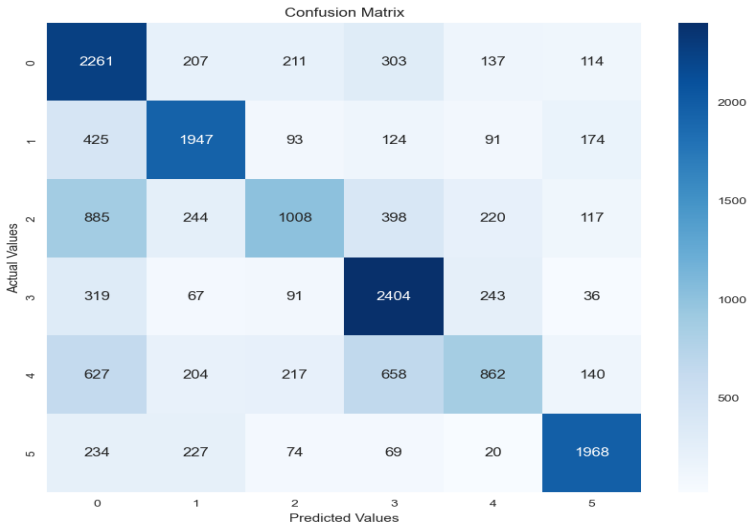


Figure 7.6: Confusion Matrix of Voting Classifier

7.1.3 For Random Forest Classifier

1. **ROC curve of Random Forest Classifier:**

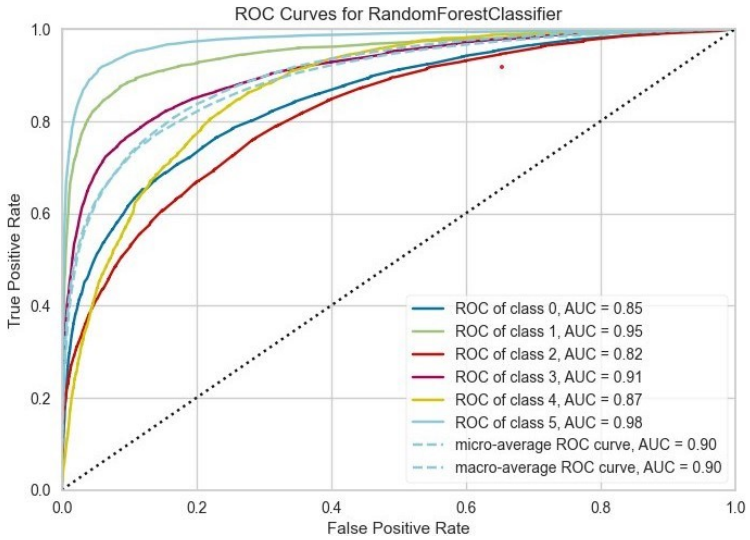


Figure 7.7: ROC-AUC curve of Random Forest Classifier

2. **Classification Report of Random Forest Classifier:**

```
Accuracy:  0.6587256027554536

Model MSE:  2.226349024110218
Model MAE:  0.7723880597014925
              precision    recall  f1-score   support

           0       0.64      0.55      0.59      3207
           1       0.83      0.77      0.80      2835
           2       0.64      0.37      0.47      2913
           3       0.69      0.73      0.71      3116
           4       0.47      0.68      0.55      2746
           5       0.77      0.88      0.82      2603

    accuracy                           0.66     17420
   macro avg       0.67      0.66      0.66     17420
weighted avg       0.67      0.66      0.65     17420
```

Figure 7.8: Classification report of Random Forest Algorithm

3. **Confusion Matrix of Random Forest Classifier:**

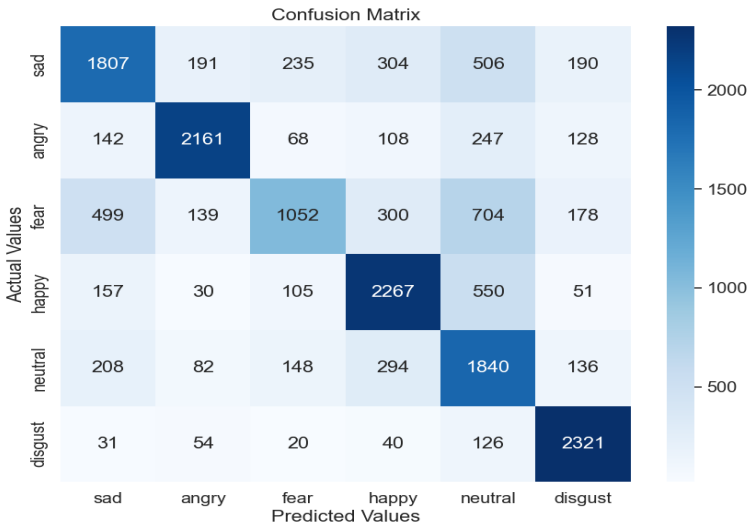


Figure 7.9: Confusion Matrix of Random Forest Classifier

7.1.4 For Support Vector Machine Classifier

1. **ROC curve of Support Vector Machine Classifier:**

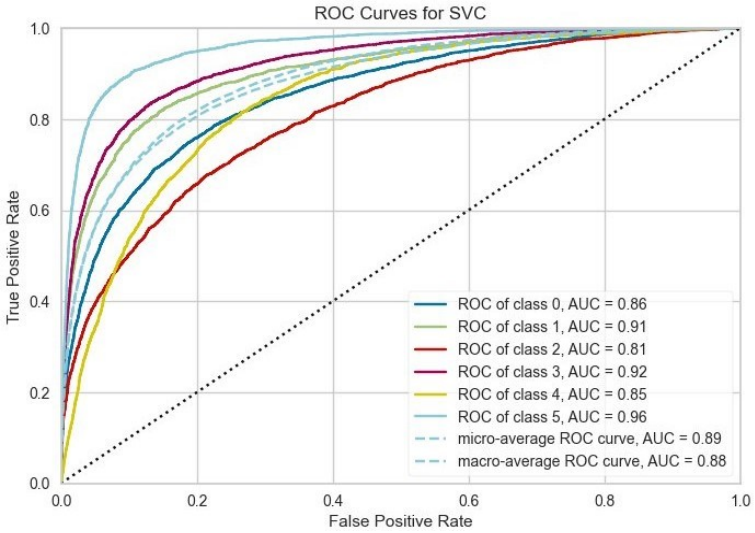


Figure 7.10: ROC-AUC curve of SVM classifier

2. **Classification Report of Support Vector Machine Algorithm:**

```
Accuracy:  0.6246268656716418

Model MSE:  2.409988518943743
Model MAE:  0.8380022962112514
              precision    recall  f1-score   support

           0       0.58      0.60      0.59      3207
           1       0.70      0.64      0.67      2835
           2       0.54      0.43      0.48      2913
           3       0.72      0.71      0.72      3116
           4       0.47      0.57      0.52      2746
           5       0.76      0.80      0.78      2603

    accuracy                           0.62     17420
   macro avg       0.63      0.63      0.63     17420
weighted avg       0.63      0.62      0.62     17420
```

Figure 7.11: Classification report of SVM Algorithm

7.2 Generative Model

We conducted experiments on a transformer model with and without considering emotion, and adjusting their hyperparameters as outlined in the table below:

S.N	Batch size	D-model	Epoch	Accuracy	Bleu-Score
1	32	128	20	0.235	0.092
2	32	256	50	0.298	0.113
3	64	128	20	0.267	0.126
4	64	256	50	0.313	0.137

Table 7.1: Evaluation results for Transformer without Emotion

S.N	Batch size	D-model	Epoch	Accuracy	Bleu-Score
1	32	128	20	0.283	0.117
2	32	256	50	0.332	0.128
3	64	128	20	0.327	0.137
4	64	256	50	0.3472	0.182

Table 7.2: Evaluation results for Transformer with Emotion

Based on the experiments conducted, we opted for the Transformer model integrated with emotion, as it demonstrated superior performance and accuracy compared to the other models tested. Consequently, we decided to employ this selected model as the backbone of our backend system.

Chapter 8

Expected Outcomes

To begin with, the user is shown the user interface A.5 . The main menu offers two choices: communicating via text or voice . The former allows interaction with the chatbot through text, while the latter facilitates input via voice. Moving on to the backend requirements, we need access to the elevenlabs and Whisper APIs for converting text to speech and voice to text, as well as our TER model and generative model.

Our backend is tasked with producing lipsync for the spoken output and transmitting it to the 3D avatar. Subsequently, the avatar reacts by displaying the corresponding emotion and gestures based on the response. Furthermore, the backend extends this process.

Then the final step is to 3D model avatar to act as natural to human being A.5 and make user experience more interactive and fun. A deployable web app using simple react three fiber and python will be made.

So, it was decided to train both TER and Generative model based on the various hyperparameter.Then, after training each of above models, these models would be compare based on evaluation metrices followed in the project.

Chapter 9

Actual Outcome

9.1 Model Evaluation

9.1.1 Evaluation of Emotion Recognition Model

We conducted model testing and evaluation using 20% of the dataset. Evaluation metrics including accuracy, precision, f1-score, recall, and others were computed, yielding the following results for each algorithm:

Metric	Value
Accuracy	65.87%
Precision	65.83%
F1-Score	65.66%
Recall	54.16%
Macro F1-Score	66%

Table 9.1: RF TER Model Evaluation

Metric	Value
Accuracy	62.46%
Precision	62.83%
F1-Score	62.67%
Recall	62.5%
Macro F1-Score	63%

Table 9.2: SVM TER Model Evaluation

Metric	Value
Accuracy	59.94%
Precision	61.5%
F1-Score	58.34%
Recall	59.67%
Macro F1-Score	59%

Table 9.3: MultinomialNB TER Model Evaluation

Metric	Value
Accuracy	66.48%
Precision	67%
F1-Score	66.34%
Recall	66.84%
Macro F1-Score	66%

Table 9.4: Voting Classifier TER Model Evaluation

Based on the provided table, the evaluation of the model based on Accuracy and Macro F1-score indicates that the Voting Classifier outperforms the other algorithms, according to our analysis.So, we used the Voting Classifier for Text Emotion Recognition.

9.2 Training of Transformer Model

1. **Training of a Transformer Model:**
   As shown in table 7.1 and 7.2 we trained the model in 20 and 50 epoches varying other hyperparameters.

```
2864/2864 [==============================] - 169s 59ms/step - loss: 0.7979 - accuracy: 0.2538
Epoch 12/20
2864/2864 [==============================] - 171s 60ms/step - loss: 0.7806 - accuracy: 0.2567
Epoch 13/20
2864/2864 [==============================] - 169s 59ms/step - loss: 0.7654 - accuracy: 0.2596
Epoch 14/20
2864/2864 [==============================] - 169s 59ms/step - loss: 0.7520 - accuracy: 0.2618
Epoch 15/20
2864/2864 [==============================] - 171s 60ms/step - loss: 0.7396 - accuracy: 0.2641
Epoch 16/20
2864/2864 [==============================] - 170s 59ms/step - loss: 0.7277 - accuracy: 0.2661
Epoch 17/20
2864/2864 [==============================] - 172s 60ms/step - loss: 0.7173 - accuracy: 0.2680
Epoch 18/20
2864/2864 [==============================] - 176s 61ms/step - loss: 0.7076 - accuracy: 0.2697
Epoch 19/20
2864/2864 [==============================] - 170s 59ms/step - loss: 0.6983 - accuracy: 0.2714
Epoch 20/20
2864/2864 [==============================] - 169s 59ms/step - loss: 0.6899 - accuracy: 0.2728
```

Figure 9.1: Training graph of Transformer Model

2. **Learning rate of a Transformer Model:**
   Regarding the learning rate, we devised a custom schedule function for the transformer model. This function adjusts the learning rate based on the provided graph.

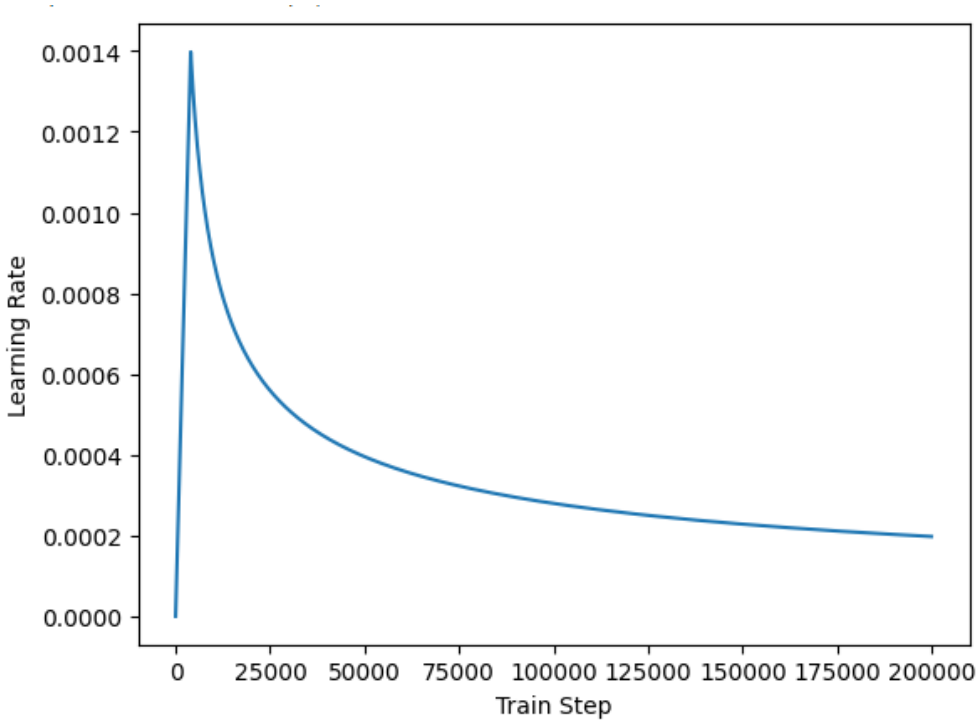


Figure 9.2: Learning rate of Transformer Model

3. **Model Selection:** As detailed in the experiment section, the Transformer model with emotion was chosen based on its performance metrics, including a BLEU score of 0.182 and an accuracy of 36%. This model was trained over 50 epochs with a batch size of 64.

4. **Evaluation of transformer model:** After selecting the model, we generated the following graph for evaluation purposes.

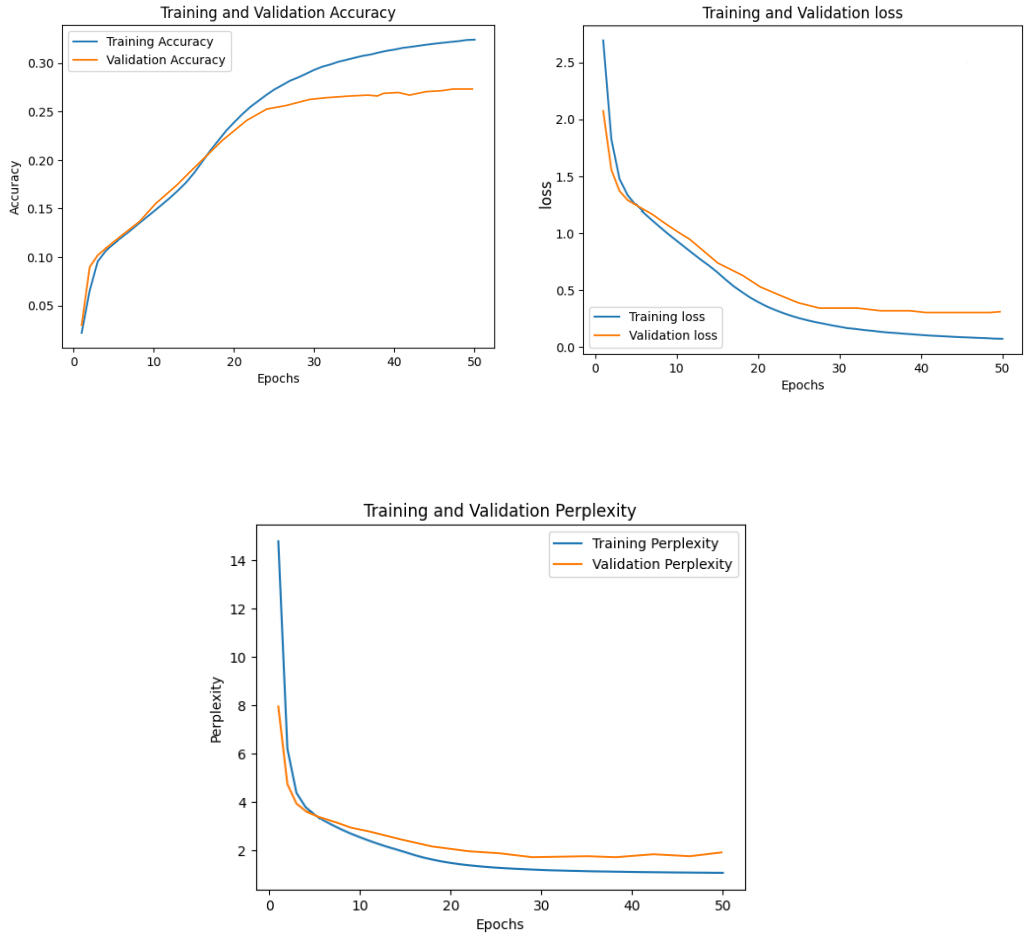


Figure 9.3: Evaluation graph of Transformer Model

Based on the graph presented above, it is evident that during the initial 20 epochs, our model displays considerable instability, likely due to training with an early learning rate of 0.001. However, at the 21st epoch, the first reduction in the learning rate occurs, resulting in a noticeable stabilization of our model. Subsequently, both the training and testing accuracy steadily improve over time, indicating effective learning from the training data.

5. **Manual Evaluation:** The generative model underwent manual testing and evaluation. An example from a sample conversation is provided in Appendix A.10. This snapshot delineates the current message, followed by the associated emotion. Evaluation primarily relied on manual assessments by team members and peers. Incorporating feedback received, additional features were integrated, and optimizations were made to enhance the model's performance.

9.3 Problem Faced

During the software development lifecycle of our project, several problems were encountered, some of which are listed as follows:

1. **During Training:** Our devices' constrained GPU capabilities hindered extensive training. When training the model on Google Colab, the GPU's computational limits restricted the number of iterations feasible at once. Consequently, the training process necessitated resuming from previous checkpoints, leading to prolonged cool-down intervals. Nevertheless, the college-provided GPU partly alleviated this challenge.
2. **Data Quality and Availability:** Securing high-quality and diverse datasets for training sentiment models, particularly Text Emotion Recognition (TER), proved challenging. Limited access to labeled data impacted the model's generalization.
3. **Real time Processing:** Making our system process emotions in real-time was a challenge. We had to balance speed and accuracy through careful optimization.

Chapter 10

Conclusion and Future Enhancements

In conclusion, our project, "Avatar Fusion: 3D Companion with Sentimental Analysis," successfully integrates sentiment models and a transformer architecture to create an empathetic 3D companion. With our TER model, we achieved high accuracy and macro F1-scores in evaluations. The transformer model generates understanding responses tailored to the user's emotional state, acting as an Emotion Recognition-based audio chatbot. Our system provides emotional support, facilitating communication and complementing conventional mental health treatments.

As per future enhancements following can be considered:

1. **Real-time Processing Optimization:** Optimize algorithms and explore hardware acceleration options to reduce latency for real-time processing.
2. **User Personalization:** Introduce personalized user profiles for tailored responses based on individual preferences and emotional nuances.
3. **User Feedback Mechanism:** Integrate a user feedback mechanism for continuous improvement based on user input and expectations.
4. **Expansion to Other Languages:** Our model only understand English language. Increase accessibility by extending language support beyond the current scope through training models on datasets in additional languages.
5. **Ethical Considerations and Bias Mitigation:** Conduct a comprehensive assessment of ethical implications and potential biases, implementing strategies for fair and inclusive behavior.

References

[1] K. Denecke, S. Vaaheesan, and A. Arulnathan, "A mental health chatbot for regulating emotions (sermo)-concept and usability test," *IEEE Transactions on Emerging Topics in Computing*, vol. 9, no. 3, pp. 1170–1182, 2020.

[2] T. Teubner, C. M. Flath, C. Weinhardt, W. van der Aalst, and O. Hinz, "Welcome to the era of chatgpt et al. the prospects of large language models," *Business & Information Systems Engineering*, vol. 65, no. 2, pp. 95–101, 2023.

[3] J. Devlin, M.-W. Chang, and K. Lee, "Google, kt, language, ai: Bert: pre-training of deep bidirectional transformers for language understanding," in *Proceedings of NAACL-HLT*, 2019, pp. 4171–4186.

[4] C.-C. Lin, A. Y. Huang, and S. J. Yang, "A review of ai-driven conversational chatbots implementation methodologies and challenges (1999–2022)," *Sustainability*, vol. 15, no. 5, p. 4012, 2023.

[5] L. Laranjo, A. G. Dunn, H. L. Tong, A. B. Kocaballi, J. Chen, R. Bashir, D. Surian, B. Gallego, F. Magrabi, A. Y. Lau *et al.*, "Conversational agents in healthcare: a systematic review," *Journal of the American Medical Informatics Association*, vol. 25, no. 9, pp. 1248–1258, 2018.

[6] A. A. Abd-Alrazaq, M. Alajlani, A. A. Alalwan, B. M. Bewick, P. Gardner, and M. Househ, "An overview of the features of chatbots in mental health: A scoping review," *International Journal of Medical Informatics*, vol. 132, p. 103978, 2019.

[7] P. A. Angga, W. E. Fachri, A. Elevanita, Suryadi, and R. D. Agushinta, "Design of chatbot with 3d avatar, voice interface, and facial expression," in *2015 International Conference on Science in Information Technology (ICSITech)*, 2015, pp. 326–330.

[8] E. Svikhnushina and P. Pu, "Key qualities of conversational chatbots–the peace model," in *26th International Conference on Intelligent User Interfaces*, 2021, pp. 520–530.

[9] Y.-T. Wan, C.-C. Chiu, K.-W. Liang, and P.-C. Chang, "Midoriko chatbot: Lstm-based emotional 3d avatar," in *2019 IEEE 8th Global Conference on Consumer Electronics (GCCE)*. IEEE, 2019, pp. 937–940.

[10] D. Das Chakladar, P. Kumar, S. Mandal, P. P. Roy, M. Iwamura, and B.-G. Kim, "3d avatar approach for continuous sign movement using speech/text," *Applied Sciences*, vol. 11, no. 8, p. 3439, 2021.

[11] S. Borsci, A. Malizia, M. Schmettow, F. Van Der Velde, G. Tariverdiyeva, D. Balaji, and A. Chamberlain, "The chatbot usability scale: the design and pilot of a usability scale for interaction with ai-based conversational agents," *Personal and Ubiquitous Computing*, vol. 26, pp. 95–119, 2022.

[12] F. A. A. J. Almahri, D. Bell, and M. Arzoky, "Applications of machine learning in education: Personas design for chatbots," in *Machine Learning Approaches for Improvising Modern Learning Systems*. IGI Global, 2021, pp. 72–113.

[13] "Svm." [Online]. Available: https://www.researchgate.net/figure/Support-Vector-Machine-visualization_fig5_332248436

[14] "Randomforest." [Online]. Available: https://towardsdatascience.com/understanding-random-forest-58381e0602d2

[15] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[16] V. Ashish, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, p. I, 2017.

[17] "Emotion dataset for emotion recognition tasks." [Online]. Available: https://www.kaggle.com/datasets/parulpandey/emotion-dataset

[18] "Emotion-dataset." [Online]. Available: https://github.com/ReetKubba/Text-Emotion-Classifier

[19] "Sentiment analysis in text." [Online]. Available: https://data.world/crowdflower/sentiment-analysis-in-text

[20] D. Demszky, D. Movshovitz-Attias, J. Ko, A. Cowen, G. Nemade, and S. Ravi, "GoEmotions: A Dataset of Fine-Grained Emotions," in *58th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2020.

[21] "Hugging face dataset for generative model." [Online]. Available: https://huggingface.co/datasets/empathetic_dialogues

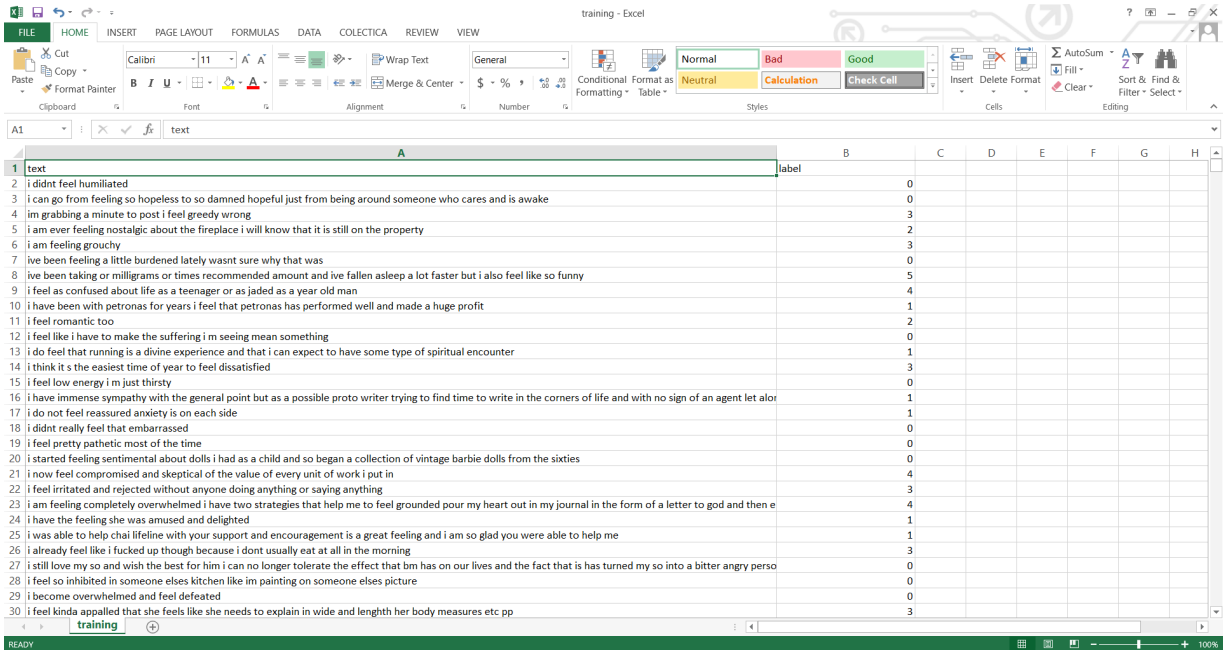
[22] H. Rashkin, E. M. Smith, M. Li, and Y.-L. Boureau, "Towards empathetic open-domain conversation models: A new benchmark and dataset," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence, Italy: Association for Computational Linguistics, Jul. 2019, pp. 5370–5381. [Online]. Available: https://aclanthology.org/P19-1534

[23] "empathic facebook dialogues." [Online]. Available: https://www.kaggle.com/datasets/atharvjairath/empathetic-dialogues-facebook-ai

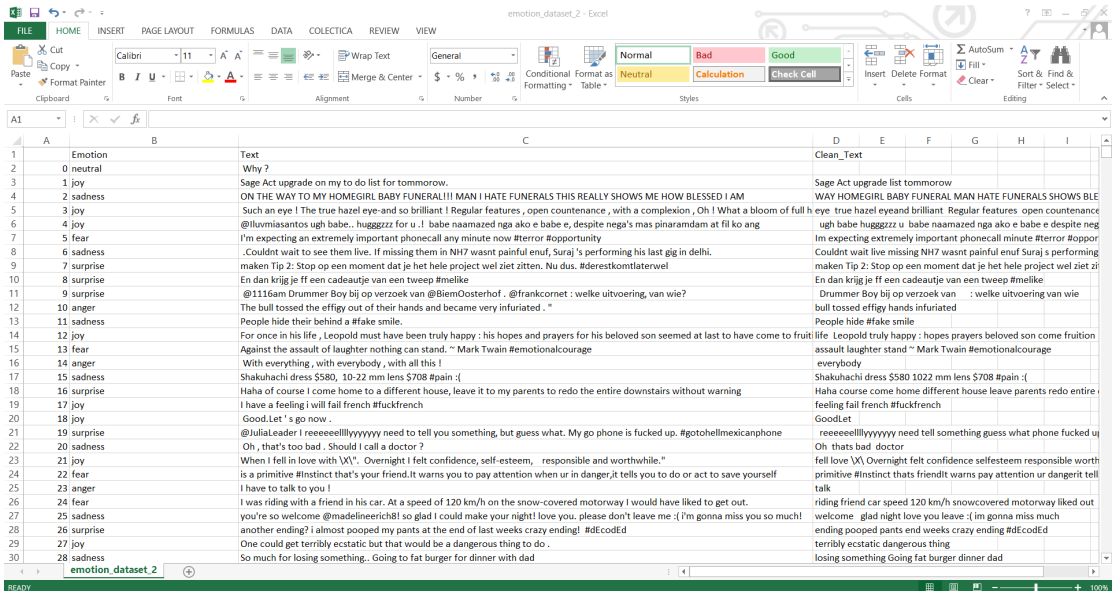
Appendix

A Snapshots

A.1 Unprocessed Dataset Sample for TER Model

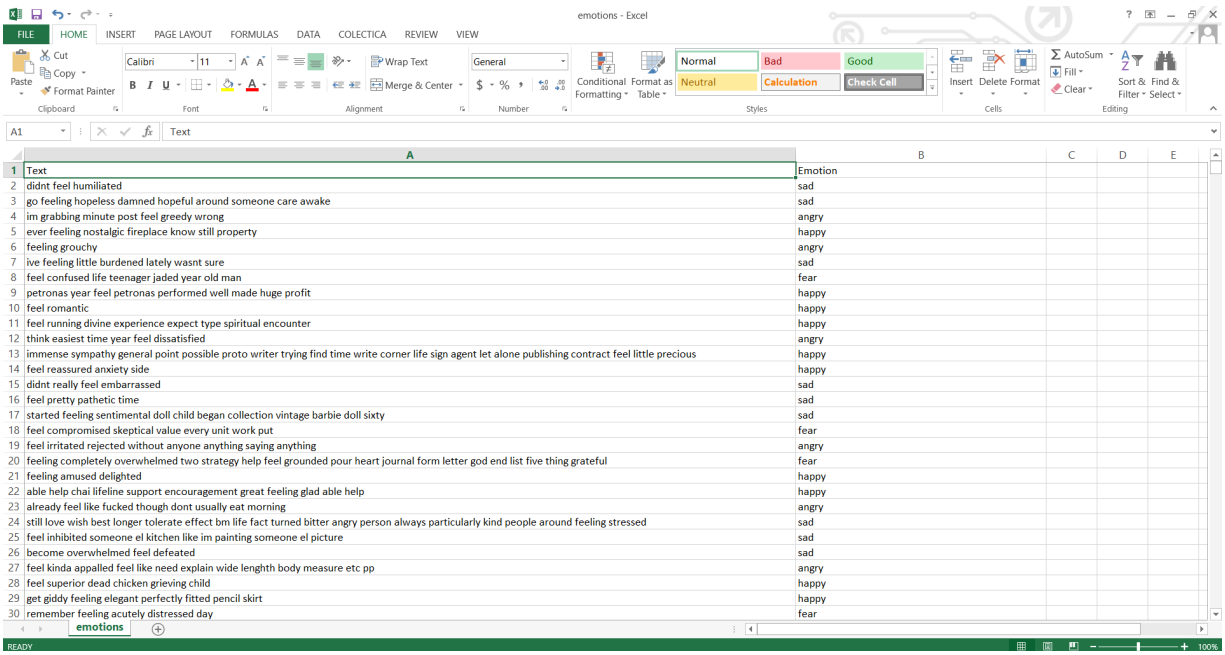


	A	B
1	text	label
2	i didnt feel humiliated	0
3	i can go from feeling so hopeless to so damned hopeful just from being around someone who cares and is awake	0
4	im grabbing a minute to post i feel greedy wrong	3
5	i am ever feeling nostalgic about the fireplace i will know that it is still on the property	2
6	i am feeling grouchy	3
7	ive been feeling a little burdened lately wasnt sure why that was	0
8	ive been taking or milligrams or times recommended amount and ive fallen asleep a lot faster but i also feel like so funny	5
9	i feel as confused about life as a teenager or as jaded as a year old man	4
10	i have been with petronas for years i feel that petronas has performed well and made a huge profit	1
11	i feel romantic too	2
12	i feel like i have to make the suffering i m seeing mean something	0
13	i do feel that running is a divine experience and that i can expect to have some type of spiritual encounter	1
14	i think it s the easiest time of year to feel dissatisfied	3
15	i feel low energy i m just thirsty	0
16	i have immense sympathy with the general point but as a possible proto writer trying to find time to write in the corners of life and with no sign of an agent let alor	1
17	i do not feel reassured anxiety is on each side	1
18	i didnt really feel that embarrassed	0
19	i feel pretty pathetic most of the time	0
20	i started feeling sentimental about dolls i had as a child and so began a collection of vintage barbie dolls from the sixties	0
21	i now feel compromised and skeptical of the value of every unit of work i put in	4
22	i feel irritated and rejected without anyone doing anything or saying anything	3
23	i am feeling completely overwhelmed i have two strategies that help me to feel grounded pour my heart out in my journal in the form of a letter to god and then e	4
24	i have the feeling she was amused and delighted	1
25	i was able to help chai lifeline with your support and encouragement is a great feeling and i am so glad you were able to help me	1
26	i already feel like i fucked up though because i dont usually eat at all in the morning	3
27	i still love my so and wish the best for him i can no longer tolerate the effect that bm has on our lives and the fact that is has turned my so into a bitter angry perso	0
28	i feel so inhibited in someone elses kitchen like im painting on someone elses picture	0
29	i become overwhelmed and feel defeated	0
30	i feel kinda appalled that she feels like she needs to explain in wide and lenghth her body measures etc pp	3



	A	B	C	D
1		Emotion	Text	Clean_Text
2	0	neutral	Why ?	
3	1	joy	Sage Act upgrade on my to do list for tommorow.	Sage Act upgrade list tommorow
4	2	sadness	ON THE WAY TO MY HOMEGIRL BABY FUNERAL!!! MAN I HATE FUNERALS THIS REALLY SHOWS ME HOW BLESSED I AM	WAY HOMEGIRL BABY FUNERAL MAN HATE FUNERALS SHOWS BLE
5	3	joy	Such an eye ! The true hazel eye-and so brilliant ! Regular features , open countenance , with a complexion , Oh ! What a bloom of full h	eye true hazel eyeand brilliant Regular features open countenance
6	4	joy	@Iluvmiasantos ugh babe.. hugggzzz for u .! babe naamazed nga ako e babe e, despite nega's mas pinaramdam at fil ko ang	ugh babe hugggzzz u babe naamazed nga ako e babe e despite neg
7	5	fear	I'm expecting an extremely important phonecall any minute now #terror #opportunity	Im expecting extremely important phonecall minute #terror #oppor
8	6	sadness	.Couldnt wait to see them live. If missing them in NH7 wasnt painful enuf, Suraj 's performing his last gig in delhi.	Couldnt wait live missing NH7 wasnt painful enuf Suraj s performing
9	7	surprise	maken Tip 2: Stop op een moment dat je het hele project wel ziet zitten. Nu dus. #derestkomtlaterwel	maken Tip 2: Stop op een moment dat je het hele project wel ziet zi
10	8	surprise	En dan krijg je ff een cadeautje van een tweep #melike	En dan krijg je ff een cadeautje van een tweep #melike
11	9	surprise	@1116am Drummer Boy bij op verzoek van @BiemOosterhof . @frankcornet : welke uitvoering, van wie?	Drummer Boy bij op verzoek van : welke uitvoering van wie
12	10	anger	The bull tossed the effigy out of their hands and became very infuriated . "	bull tossed effigy hands infuriated
13	11	sadness	People hide their behind a #fake smile.	People hide #fake smile
14	12	joy	For once in his life , Leopold must have been truly happy : his hopes and prayers for his beloved son seemed at last to have come to fruit	life Leopold truly happy : hopes prayers beloved son come fruition
15	13	fear	Against the assault of laughter nothing can stand. ~ Mark Twain #emotionalcourage	assault laughter stand ~ Mark Twain #emotionalcourage
16	14	anger	With everything , with everybody , with all this !	everybody
17	15	sadness	Shakuhachi dress $580, 10-22 mm lens $708 #pain :(	Shakuhachi dress $580 1022 mm lens $708 #pain :(
18	16	surprise	Haha of course I come home to a different house, leave it to my parents to redo the entire downstairs without warning	Haha course come home different house leave parents redo entire
19	17	joy	I have a feeling i will fail french #fuckfrench	feeling fail french #fuckfrench
20	18	joy	Good.Let ' s go now .	GoodLet
21	19	surprise	@JuliaLeader I reeeeeellllyyyyyyy need to tell you something, but guess what. My go phone is fucked up. #gotohellmexicanphone	reeeeeellllyyyyyyy need tell something guess what phone fucked u
22	20	sadness	Oh , that's too bad . Should I call a doctor ?	Oh thats bad doctor
23	21	joy	When I fell in love with \X\". Overnight I felt confidence, self-esteem, responsible and worthwhile."	fell love \X\ Overnight felt confidence selfesteem responsible worth
24	22	fear	is a primitive #Instinct that's your friend.It warns you to pay attention when ur in danger,it tells you to do or act to save yourself	primitive #Instinct thats friendIt warns pay attention ur dangerit tell
25	23	anger	I have to talk to you !	talk
26	24	fear	I was riding with a friend in his car. At a speed of 120 km/h on the snow-covered motorway I would have liked to get out.	riding friend car speed 120 km/h snowcovered motorway liked out
27	25	sadness	you're so welcome @madelinerich8! so glad I could make your night! love you. please don't leave me :(i'm gonna miss you so much!	welcome glad night love you leave :(im gonna miss much
28	26	surprise	another ending? i almost pooped my pants at the end of last weeks crazy ending! #dEcodEd	ending pooped pants end weeks crazy ending #dEcodEd
29	27	joy	One could get terribly ecstatic but that would be a dangerous thing to do .	terribly ecstatic dangerous thing
30	28	sadness	So much for losing something.. Going to fat burger for dinner with dad	losing something Going fat burger dinner dad

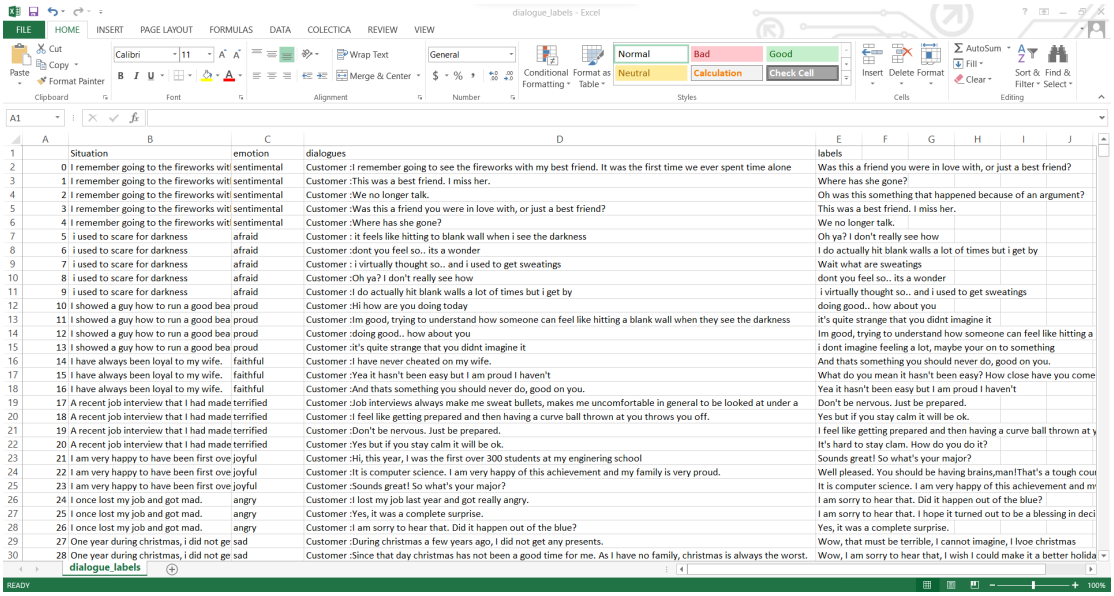
A.2 Processed Dataset Sample for TER Model



	A	B
1	Text	Emotion
2	didnt feel humiliated	sad
3	go feeling hopeless damned hopeful around someone care awake	sad
4	im grabbing minute post feel greedy wrong	angry
5	ever feeling nostalgic fireplace know still property	happy
6	feeling grouchy	angry
7	ive feeling little burdened lately wasnt sure	sad
8	feel confused life teenager jaded year old man	fear
9	petronas year feel petronas performed well made huge profit	happy
10	feel romantic	happy
11	feel running divine experience expect type spiritual encounter	happy
12	think easiest time year feel dissatisfied	angry
13	immense sympathy general point possible proto writer trying find time write corner life sign agent let alone publishing contract feel little precious	happy
14	feel reassured anxiety side	happy
15	didnt really feel embarrassed	sad
16	feel pretty pathetic time	sad
17	started feeling sentimental doll child began collection vintage barbie doll sixty	sad
18	feel compromised skeptical value every unit work put	fear
19	feel irritated rejected without anyone anything saying anything	angry
20	feeling completely overwhelmed two strategy help feel grounded pour heart journal form letter god end list five thing grateful	fear
21	feeling amused delighted	happy
22	able help chai lifeline support encouragement great feeling glad able help	happy
23	already feel like fucked though dont usually eat morning	angry
24	still love wish best longer tolerate effect bm life fact turned bitter angry person always particularly kind people around feeling stressed	sad
25	feel inhibited someone el kitchen like im painting someone el picture	sad
26	become overwhelmed feel defeated	sad
27	feel kinda appalled feel like need explain wide lenghth body measure etc pp	angry
28	feel superior dead chicken grieving child	happy
29	get giddy feeling elegant perfectly fitted pencil skirt	happy
30	remember feeling acutely distressed day	fear

A.3 Unprocessed Dataset Sample for Generative Model

conv_id	utterance_idx	context	prompt	speaker_ic	utterance	selfeval	tags	junk
hit:0_conv:1	1	sentimental	I remember going to the firework:	1	I remember going to see the fireworks with my best friend. It	5|5|5_2|2|5		
hit:0_conv:1	2	sentimental	I remember going to the firework:	0	Was this a friend you were in love with_comma_ or just a be	5|5|5_2|2|5		
hit:0_conv:1	3	sentimental	I remember going to the firework:	1	This was a best friend. I miss her.	5|5|5_2|2|5		
hit:0_conv:1	4	sentimental	I remember going to the firework:	0	Where has she gone?	5|5|5_2|2|5		
hit:0_conv:1	5	sentimental	I remember going to the firework:	1	We no longer talk.	5|5|5_2|2|5		
hit:0_conv:1	6	sentimental	I remember going to the firework:	0	Oh was this something that happened because of an argume	5|5|5_2|2|5		
hit:1_conv:2	1	afraid	i used to scare for darkness	2	it feels like hitting to blank wall when i see the darkness	4|3|4_3|5|5		
hit:1_conv:2	2	afraid	i used to scare for darkness	3	Oh ya? I don't really see how	4|3|4_3|5|5		
hit:1_conv:2	3	afraid	i used to scare for darkness	2	dont you feel so.. its a wonder	4|3|4_3|5|5		
hit:1_conv:2	4	afraid	i used to scare for darkness	3	I do actually hit blank walls a lot of times but i get by	4|3|4_3|5|5		
hit:1_conv:2	5	afraid	i used to scare for darkness	2	i virtually thought so.. and i used to get sweatings	4|3|4_3|5|5		
hit:1_conv:2	6	afraid	i used to scare for darkness	3	Wait what are sweatings	4|3|4_3|5|5		
hit:1_conv:3	1	proud	I showed a guy how to run a good	3	Hi how are you doing today	3|5|5_4|3|4	<HI>	
hit:1_conv:3	2	proud	I showed a guy how to run a good	2	doing good.. how about you	3|5|5_4|3|4		
hit:1_conv:3	3	proud	I showed a guy how to run a good	3	Im good_comma_ trying to understand how someone can fe	3|5|5_4|3|4		
hit:1_conv:3	4	proud	I showed a guy how to run a good	2	it's quite strange that you didnt imagine it	3|5|5_4|3|4		
hit:1_conv:3	5	proud	I showed a guy how to run a good	3	i dont imagine feeling a lot_comma_ maybe your on to some	3|5|5_4|3|4		
hit:2_conv:4	1	faithful	I have always been loyal to my wi	4	I have never cheated on my wife.	3|3|5_2|4|4		
hit:2_conv:4	2	faithful	I have always been loyal to my wi	5	And thats something you should never do_comma_ good on	3|3|5_2|4|4		
hit:2_conv:4	3	faithful	I have always been loyal to my wi	4	Yea it hasn't been easy but I am proud I haven't	3|3|5_2|4|4		
hit:2_conv:4	4	faithful	I have always been loyal to my wi	5	What do you mean it hasn't been easy? How close have you	3|3|5_2|4|4		
hit:2_conv:5	1	terrified	A recent job interview that I had r	5	Job interviews always make me sweat bullets_comma_ make	2|4|4_3|3|5		

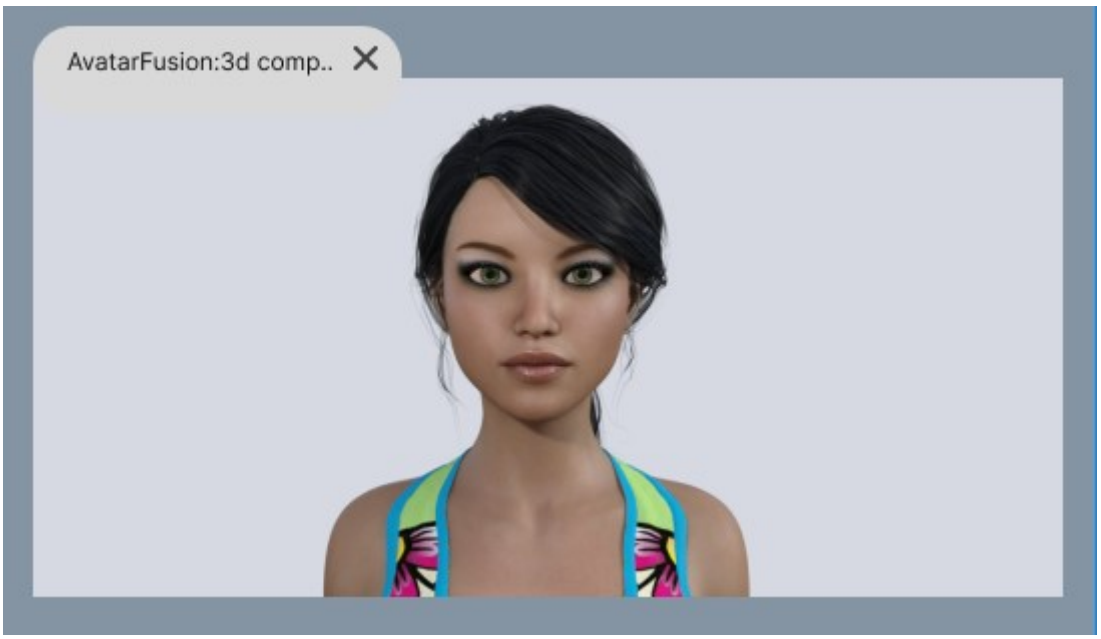


		Situation	emotion	dialogues	labels
2	0	I remember going to the fireworks wit	sentimental	Customer :I remember going to see the fireworks with my best friend. It was the first time we ever spent time alone	Was this a friend you were in love with, or just a best friend?
3	1	I remember going to the fireworks wit	sentimental	Customer :This was a best friend. I miss her.	Where has she gone?
4	2	I remember going to the fireworks wit	sentimental	Customer :We no longer talk.	Oh was this something that happened because of an argument?
5	3	I remember going to the fireworks wit	sentimental	Customer :Was this a friend you were in love with, or just a best friend?	This was a best friend. I miss her.
6	4	I remember going to the fireworks wit	sentimental	Customer :Where has she gone?	We no longer talk.
7	5	i used to scare for darkness	afraid	Customer : it feels like hitting to blank wall when i see the darkness	Oh ya? I don't really see how
8	6	i used to scare for darkness	afraid	Customer :dont you feel so.. its a wonder	I do actually hit blank walls a lot of times but i get by
9	7	i used to scare for darkness	afraid	Customer : i virtually thought so.. and i used to get sweatings	Wait what are sweatings
10	8	i used to scare for darkness	afraid	Customer :Oh ya? I don't really see how	dont you feel so.. its a wonder
11	9	i used to scare for darkness	afraid	Customer :I do actually hit blank walls a lot of times but i get by	i virtually thought so.. and i used to get sweatings
12	10	I showed a guy how to run a good bea	proud	Customer :Hi how are you doing today	doing good.. how about you
13	11	I showed a guy how to run a good bea	proud	Customer :Im good, trying to understand how someone can feel like hitting a blank wall when they see the darkness	it's quite strange that you didnt imagine it
14	12	I showed a guy how to run a good bea	proud	Customer :doing good.. how about you	Im good, trying to understand how someone can feel like hitting a
15	13	I showed a guy how to run a good bea	proud	Customer :it's quite strange that you didnt imagine it	i dont imagine feeling a lot, maybe your on to something
16	14	I have always been loyal to my wife.	faithful	Customer :I have never cheated on my wife.	And thats something you should never do, good on you.
17	15	I have always been loyal to my wife.	faithful	Customer :Yea it hasn't been easy but I am proud I haven't	What do you mean it hasn't been easy? How close have you come
18	16	I have always been loyal to my wife.	faithful	Customer :And thats something you should never do, good on you.	Yea it hasn't been easy but I am proud I haven't
19	17	A recent job interview that I had made	terrified	Customer :Job interviews always make me sweat bullets, makes me uncomfortable in general to be looked at under a	Don't be nervous. Just be prepared.
20	18	A recent job interview that I had made	terrified	Customer :I feel like getting prepared and then having a curve ball thrown at you throws you off.	Yes but if you stay calm it will be ok.
21	19	A recent job interview that I had made	terrified	Customer :Don't be nervous. Just be prepared.	I feel like getting prepared and then having a curve ball thrown at y
22	20	A recent job interview that I had made	terrified	Customer :Yes but if you stay calm it will be ok.	It's hard to stay clam. How do you do it?
23	21	I am very happy to have been first ove	joyful	Customer :Hi, this year, I was the first over 300 students at my enginering school	Sounds great! So what's your major?
24	22	I am very happy to have been first ove	joyful	Customer :It is computer science. I am very happy of this achievement and my family is very proud.	Well pleased. You should be having brains,man!That's a tough cou
25	23	I am very happy to have been first ove	joyful	Customer :Sounds great! So what's your major?	It is computer science. I am very happy of this achievement and m
26	24	I once lost my job and got mad.	angry	Customer :I lost my job last year and got really angry.	I am sorry to hear that. Did it happen out of the blue?
27	25	I once lost my job and got mad.	angry	Customer :Yes, it was a complete surprise.	I am sorry to hear that. I hope it turned out to be a blessing in deci
28	26	I once lost my job and got mad.	angry	Customer :I am sorry to hear that. Did it happen out of the blue?	Yes, it was a complete surprise.
29	27	One year during christmas, i did not ge	sad	Customer :During christmas a few years ago, I did not get any presents.	Wow, that must be terrible, I cannot imagine, I lvoe christmas
30	28	One year during christmas, i did not ge	sad	Customer :Since that day christmas has not been a good time for me. As I have no family, christmas is always the worst.	Wow, I am sorry to hear that, I wish I could make it a better holida

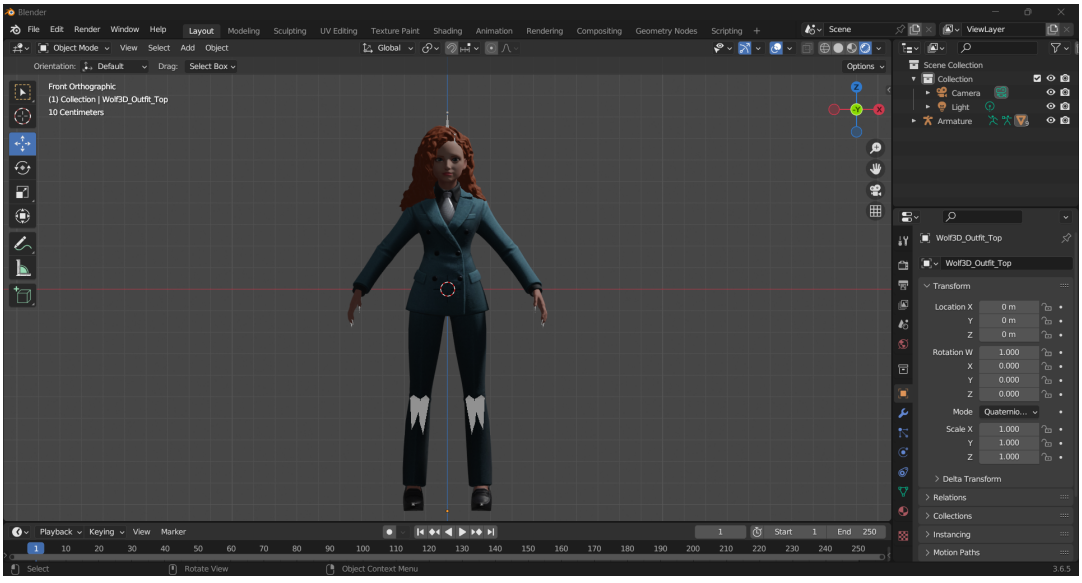
A.4 Processed Dataset Sample for Generative Model

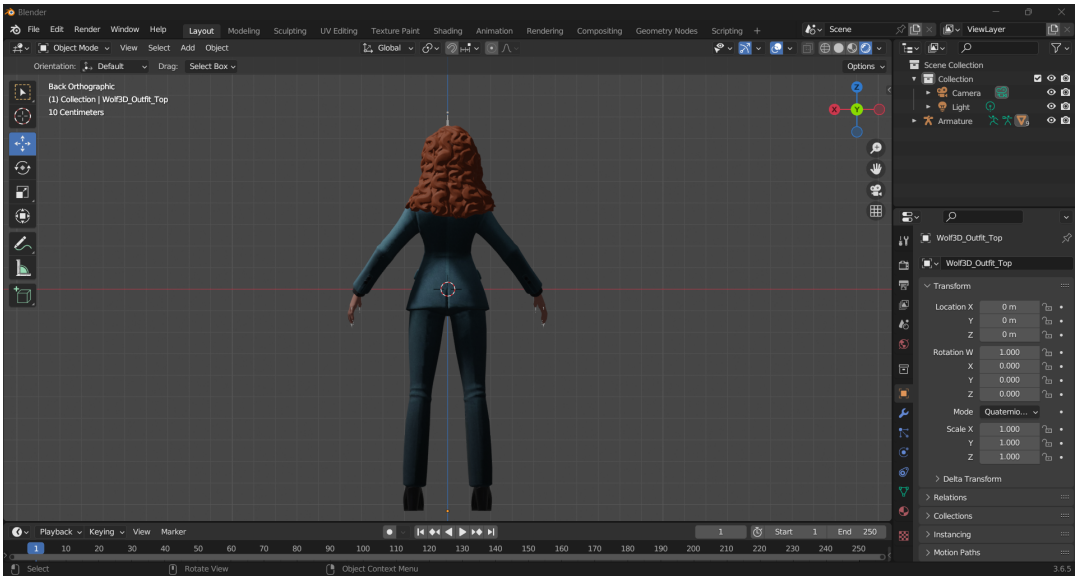
	A	B	C	D
1	Messege	Reply	Emotion	
2	i remember going to see the fireworks with my best friend . it was the first time we ever	Was this a friend you were in love with_comma_ or just a best friend?	joy	
3	was this a friend you were in love with comma or just a best friend ?	This was a best friend. I miss her.	sadness	
4	this was a best friend . i miss her .	Where has she gone?	joy	
5	where has she gone ?	We no longer talk.	fear	
6	we no longer talk .	Oh was this something that happened because of an argument?	joy	
7	it feels like hitting to blank wall when i see the darkness	Oh ya? I don't really see how	joy	
8	oh ya ? i do not really see how	dont you feel so.. its a wonder	angry	
9	dont you feel so . . its a wonder	I do actually hit blank walls a lot of times but i get by	joy	
10	i do actually hit blank walls a lot of times but i get by	i virtually thought so.. and i used to get sweatings	neutral	
11	i virtually thought so . . and i used to get sweatings	Wait what are sweatings	joy	
12	hi how are you doing today	doing good.. how about you	neutral	
13	doing good . . how about you	Im good_comma_ trying to understand how someone can feel like hitting a blank wall when th	joy	
14	im good comma trying to understand how someone can feel like hitting a blank wall wh	it's quite strange that you didnt imagine it	angry	
15	it is quite strange that you didnt imagine it	i dont imagine feeling a lot_comma_ maybe your on to something	sadness	
16	i have never cheated on my wife .	And thats something you should never do_comma_ good on you.	joy	
17	and thats something you should never do comma good on you .	Yea it hasn't been easy but I am proud I haven't	joy	
18	yea it has not been easy but i am proud i have not	What do you mean it hasn't been easy? How close have you come to cheating?	angry	
19	job interviews always make me sweat bullets comma makes me uncomfortable in gene	Don't be nervous. Just be prepared.	joy	
20	do not be nervous . just be prepared .	I feel like getting prepared and then having a curve ball thrown at you throws you off.	joy	
21	i feel like getting prepared and then having a curve ball thrown at you throws you off .	Yes but if you stay calm it will be ok.	angry	
22	yes but if you stay calm it will be ok .	It's hard to stay clam. How do you do it?	joy	
23	hi comma this year comma i was the first over 300 students at my enginering school	Sounds great! So what's your major?	neutral	
24	sounds great ! so that is your major ?	It is computer science. I am very happy of this achievement and my family is very proud.	joy	
25	it is computer science . i am very happy of this achievement and my family is very proud	Well pleased. You should be having brains_comma_man!That's a tough course_comma_i hear	joy	
26	i lost my job last year and got really angry .	I am sorry to hear that. Did it happen out of the blue?	joy	
27	i am sorry to hear that . did it happen out of the blue ?	Yes_comma_ it was a complete surprise.	joy	
28	yes comma it was a complete surprise .	I am sorry to hear that. I hope it turned out to be a blessing in decise.	joy	
29	during christmas a few years ago comma i did not get any presents .	Wow_comma_ that must be terrible_comma_ I cannot imagine_comma_ I lvoe christmas	joy	
30	wow comma that must be terrible comma i cannot imagine comma i lvoe christmas	Since that day christmas has not been a good time for me. As I have no family_comma_ christ	neutral	

A.5 Expected Outcome

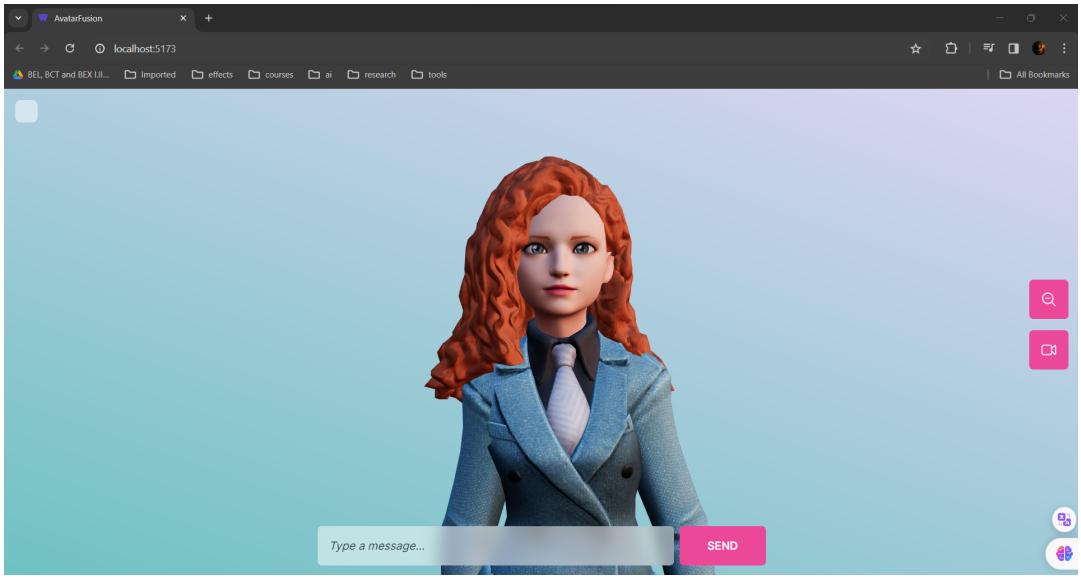


A.6 3D Avatar Development





A.7 User Interface Development



A.8 Backend Working

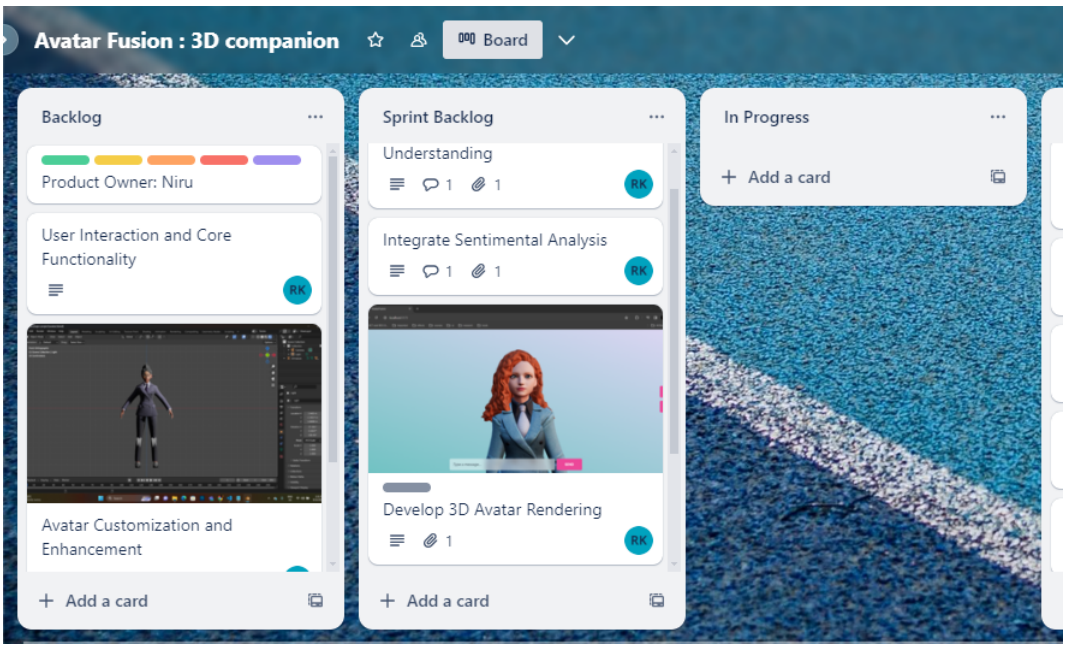
```
Microsoft Windows [Version 10.0.22621.3155]
(c) Microsoft Corporation. All rights reserved.

B:\Avatarfusion\backend-part>uvicorn API1:app --host 0.0.0.0 --port 3000 --reload
INFO:     Will watch for changes in these directories: ['B:\\Avatarfusion\\backend-part']
INFO:     Uvicorn running on http://0.0.0.0:3000 (Press CTRL+C to quit)
INFO:     Started reloader process [15044] using StatReload
neutral
2024-03-07 10:01:51.116846: I tensorflow/core/platform/cpu_feature_guard.cc:182] This TensorFlow binary is optimized to use available CPU instructions in pe
rformance-critical operations.
To enable the following instructions: SSE SSE2 SSE3 SSE4.1 SSE4.2 AVX AVX2 AVX_VNNI FMA, in other operations, rebuild TensorFlow with the appropriate compil
er flags.
INFO:     Started server process [23308]
INFO:     Waiting for application startup.
INFO:     Application startup complete.
INFO:     127.0.0.1:64773 - "OPTIONS /chat HTTP/1.1" 200 OK
hello
****reply: hi , how are you ?
****mytext: hello
reply_motion: neutral
ffmpeg version 6.1.1-full_build-www.gyan.dev Copyright (c) 2000-2023 the FFmpeg developers
  built with gcc 12.2.0 (Rev10, Built by MSYS2 project)
  configuration: --enable-gpl --enable-version3 --enable-static --pkg-config=pkgconf --disable-w32threads --disable-autodetect --enable-fontconfig --enable-
iconv --enable-gnutls --enable-libxml2 --enable-gmp --enable-bzlib --enable-lzma --enable-libsnappy --enable-zlib --enable-librist --enable-libsrt --enable-
libssh --enable-libzmq --enable-avisynth --enable-libbluray --enable-libcaca --enable-sdl2 --enable-libaribb24 --enable-libaribcaption --enable-libdav1d --e
nable-libdavs2 --enable-libuavs3d --enable-libzvbi --enable-librav1e --enable-libsvtav1 --enable-libwebp --enable-libx264 --enable-libx265 --enable-libxavs2
 --enable-libxvid --enable-libaom --enable-libjxl --enable-libopenjpeg --enable-libvpx --enable-mediafoundation --enable-libass --enable-frei0r --enable-lib
freetype --enable-libfribidi --enable-libharfbuzz --enable-liblensfun --enable-libvidstab --enable-libvmaf --enable-libzimg --enable-amf --enable-cuda-llvm
--enable-cuvid --enable-ffnvcodec --enable-nvdec --enable-nvenc --enable-dxva2 --enable-d3d11va --enable-libvpl --enable-libshaderc --enable-vulkan --enable
-libplacebo --enable-opencl --enable-libcdio --enable-libgme --enable-libmodplug --enable-libopenmpt --enable-libopencore-amrwb --enable-libmp3lame --enable
-libshine --enable-libtheora --enable-libtwolame --enable-libvo-amrwbenc --enable-libcodec2 --enable-libilbc --enable-libgsm --enable-libopencore-amrnb --en
able-libopus --enable-libspeex --enable-libvorbis --enable-ladspa --enable-libbs2b --enable-libflite --enable-libmysofa --enable-librubberband --enable-libs
oxr --enable-chromaprint
  libavutil      58. 29.100 / 58. 29.100
  libavcodec     60. 31.102 / 60. 31.102
  libavformat    60. 16.100 / 60. 16.100
  libavdevice    60.  3.100 / 60.  3.100
  libavfilter     9. 12.100 /  9. 12.100
  libswscale      7.  5.100 /  7.  5.100
  libswresample   4. 12.100 /  4. 12.100
  libpostproc    57.  3.100 / 57.  3.100
```

```
[mp3 @ 0000025d94578400] Estimating duration from bitrate, this may be inaccurate
Input #0, mp3, from 'audios/ea1cc42a-790f-493b-8723-27689b58c233.mp3':
  Duration: 00:00:02.14, start: 0.000000, bitrate: 32 kb/s
  Stream #0:0: Audio: mp3, 24000 Hz, mono, fltp, 32 kb/s
Stream mapping:
  Stream #0:0 -> #0:0 (mp3 (mp3float) -> pcm_s16le (native))
Press [q] to stop, [?] for help
Output #0, wav, to 'audios/ea1cc42a-790f-493b-8723-27689b58c233.wav':
  Metadata:
    ISFT            : Lavf60.16.100
  Stream #0:0: Audio: pcm_s16le ([1][0][0][0] / 0x0001), 24000 Hz, mono, s16, 384 kb/s
    Metadata:
      encoder         : Lavc60.31.102 pcm_s16le
[out#0/wav @ 0000025d945abe40] video:0kB audio:100kB subtitle:0kB other streams:0kB global headers:0kB muxing overhead: 0.076077%
size=     100kB time=00:00:02.11 bitrate= 388.7kbits/s speed= 150x
Generating lip sync data for audios/ea1cc42a-790f-493b-8723-27689b58c233.wav.
Progress: [####################] 100%
Done.
filename ea1cc42a-790f-493b-8723-27689b58c233
lipsync: {
  "metadata": {
    "soundFile": "B:\\Avatarfusion\\backend-part\\audios\\ea1cc42a-790f-493b-8723-27689b58c233.wav",
    "duration": 2.13
  },
  "mouthCues": [
    { "start": 0.00, "end": 0.05, "value": "X" },
    { "start": 0.05, "end": 0.10, "value": "B" },
    { "start": 0.10, "end": 0.21, "value": "G" },
    { "start": 0.21, "end": 0.35, "value": "C" },
    { "start": 0.35, "end": 0.42, "value": "B" },
    { "start": 0.42, "end": 0.49, "value": "C" },
    { "start": 0.49, "end": 0.63, "value": "B" },
    { "start": 0.63, "end": 0.91, "value": "X" },
    { "start": 0.91, "end": 0.96, "value": "B" },
    { "start": 0.96, "end": 1.01, "value": "G" },
    { "start": 1.01, "end": 1.08, "value": "C" },
    { "start": 1.08, "end": 1.22, "value": "E" },
    { "start": 1.22, "end": 1.29, "value": "B" },
    { "start": 1.29, "end": 1.57, "value": "C" },
    { "start": 1.57, "end": 1.71, "value": "B" },
    { "start": 1.71, "end": 2.13, "value": "X" }
```

```
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA' lipsync='{\n  "metadata": {\n    "soundFile": "B:\\\\Avatarfusion\\\\backend-part\\\\audios\\\\ea1cc42a-790
f-493b-8723-27689b58c233.wav",\n    "duration": 2.13\n  },\n  "mouthCues": [\n    { "start": 0.00, "end": 0.05, "value": "X" },\n    { "start": 0.05, "end":
 0.10, "value": "B" },\n    { "start": 0.10, "end": 0.21, "value": "G" },\n    { "start": 0.21, "end": 0.35, "value": "C" },\n    { "start": 0.35, "end": 0.
42, "value": "B" },\n    { "start": 0.42, "end": 0.49, "value": "C" },\n    { "start": 0.49, "end": 0.63, "value": "B" },\n    { "start": 0.63, "end": 0.91,
 "value": "X" },\n    { "start": 0.91, "end": 0.96, "value": "B" },\n    { "start": 0.96, "end": 1.01, "value": "G" },\n    { "start": 1.01, "end": 1.08, "v
alue": "C" },\n    { "start": 1.08, "end": 1.22, "value": "E" },\n    { "start": 1.22, "end": 1.29, "value": "B" },\n    { "start": 1.29, "end": 1.57, "valu
e": "C" },\n    { "start": 1.57, "end": 1.71, "value": "B" },\n    { "start": 1.71, "end": 2.13, "value": "X" }\n  ]\n}\n' facialExpression='neutral' animat
ion='neutral'
INFO:     127.0.0.1:64773 - "POST /chat HTTP/1.1" 200 OK
```

A.9 Use of Trello Application for Project Management



A.10 Transformer Model Evaluation

```
predict("i am happy",'neutral')
```

```
'I bet! It has to feel like a huge accomplishment.'
```

```
predict(" how are you doing?",'neutral')
```

```
'Got into a motorcycle accident and is now paraplegic.
```

Figure 10.1: Training With emotion

```
output = predict("hi, how are you?")
```

```
Input: hi, how are you?
Output: I am not sure yet. I was really nervous
```

```
output = predict("how are you doing?")
```

```
Input: how are you doing?
Output: I am doing well. I will be spending time on my birthday until we got the going.
```

Figure 10.2: Training Without emotion