AI-POWERED THESPIAN: DEVELOPING AN INNOVATIVE CONVERSATIONAL ENGINE

A DISSERTATION SUBMITTED TO MANCHESTER METROPOLITAN UNIVERSITY

FOR THE DEGREE OF MASTER OF SCIENCE

IN THE FACULTY OF SCIENCE AND ENGINEERING

2023

By

Richard Ademola Ogundele

Department of Computing and Mathematics

Supervisors:

Prof YongHong Peng, Dr David Jackson

Contents

List of Figures

List of Tables

Abstract

This dissertation project delves into the burgeoning realm of conversational AI, with a specific focus on its application in creative AI and storytelling. It centres around the development of a conversational theatrical engine and explores the latest advancements in conversational AI architecture. While acknowledging the resource-intensive nature of building conversational engines, including meticulous preparation, user testing, and technical expertise, this study underscores the potential of conversational agents as efficient learning tools that offer both accessibility and privacy assurances. The project showcases the development of a mobile application utilizing the waterfall methodology and harnessing the capabilities of the latest GPT model, GPT-4, boasting an impressive 1.76 trillion parameters. This development leverages Python programming and the FastAPI framework for the backend, while the frontend is crafted using the React Native framework. Rigorous testing, conducted through Postman and feedback collection from diverse users, has revealed insightful statistics. Notably, 69.7% of users had never previously encountered a similar storytelling app, indicating a unique offering. Moreover, a substantial 67.8% of users expressed comfort in interacting with an AI tool, with 16.1% holding a neutral stance. The application itself is designed for creative AI, functioning as a thespian, and facilitating interactive storytelling experiences. User feedback revealed diverse usage patterns, with 39.4% utilizing it for learning and seeking answers, 36.4% for storytelling purposes, and 24.2% employing the app to generate scripts for acting. This project not only highlights the potential of conversational AI in creative contexts but also provides valuable insights into user perceptions and adoption, paving the way for future advancements in the field.

Declaration

No part of this project has been submitted in support of an application for any other degree or qualification at this or any other institute of learning. Apart from those parts of the project containing citations to the work of others, this project is my own unaided work. It has been undertaken in accordance with the University research ethics standards, by the terms of permit number 57156.

Signed

…….

Acknowledgments

I want to thank God for trusting me to represent him in artificial intelligence. I also appreciate my parents, Mr. and Mrs. Alexander Ogundele, for their full sponsorship of my master's programme academic pursuit.

My special appreciation goes to Professor YongHong Peng and Dr. David Jackson for holding my hands and giving me clarity at every junction of confusion in this project. Thank you for your constant support and guidance towards the success of the project.

I want to express my appreciation to Manchester Metropolitan University for providing me with the personal laptop computer I needed to begin my project with the crisis support team. I also want to thank my classmates for their assistance in providing me with the answers to my project-related queries.

Abbreviations

AI……..…… Artificial Intelligence

NL…….……Natural Language Processing

NLU……Natural Language Understanding

LLM…….…….Large Language Models

CA……...…Conversational Agents

LSTM……..…Long Short-Term Memory

BERT……...……Bidirectional Encoder Representations from Transformers

GPT……..……Generative Pre-Trained Transformers

BART……Bidirectional and Auto-Regressive Transformer

Chapter One - Introduction

1.1 Background of the Study

Conversational AI agents have become widely used in various domains such as business, healthcare, e-commerce, customer service, fintech, tourism, and education. This widespread adoption can be attributed to significant advancements in machine learning, natural language processing, and Generative AI techniques, which enable the development of highly accurate models. Given the heavy reliance on conversations with natural languages in daily operations within these domains, conversational AI agents are seen as a natural fit. The increasing demand for conversational AI has sparked a rapid pace of research and development, resulting in daily innovations. However, this surge of interest in the field has also highlighted exciting yet volatile research opportunities. (Kulkarni, P. et al, 2020)

Conversational agents are agents that can hold human-like conversations with humans with the use of natural language processing and machine learning to analyse user's input and generate responses based on the context of the conversation. The user experience will be enhanced, and dialogues will be more engaging when the agent can give anecdotes, stories, and experiences related to its persona. This software engine has the potential to be used as a collaborative storytelling tool, author story scripts, and analyse monologues, and enable it users to actively participate to the story's development. Players and viewers can engage in dialogue with the conversational agent powered by AI to help shape and advance the plot. The interactive story can be advanced by users having back-and-forth discussions, asking questions, offering prompts, or making decisions.

1.2 Research Objectives

The Project Aim is to be able to provide a platform to demonstrate a system for creative AI and storytelling and hold real-time conversations with users leveraging on the latest GPT-style API which is to also be used in a variety of settings and applications, including those related to health, art, and entertainment.

The Product is a Mobile based application that leverages the latest GPT API. It will be able to interact with users through text and speech to create interactive stories and dialogues.

To achieve the aims identified Objectives need to be met:

- Develop a software system that effectively manages the entire process of conversation process and can perform speech to speech operation.
- Effective natural language prompts for GPT that portrays character and narrative while also testing for effectiveness in terms of coherence and latency.
- Ability to have the potential to store conversational exchanges for future fine-tuning purposes.
- To evaluate the products by stakeholders and compile their response.

1.3 Problem Statement

Ensuring that the conversational engine generate responses that are natural and consistent might be a significant challenge and it can be difficult to manage nuances, humour, ambiguity and keeping a consistent character depiction.

Generative AI and conversational engine systems can unintentionally pick up prejudices from training data, which could result in responses that are biased or unsuitable. A key component of this AI development creating plans to reduce biases and keep the system's replies fair which might be a challenge.

1.4 Justification of the Study

The work is important because it develops conversational AI, improves user experiences, and has widespread industry implications. It advances the understanding of natural language, entertainment innovation, theatre arts, human-AI cooperation, and AI community research. It also raises questions about the ethical and societal implications of conversational AI.

1.5 Scope of the Study

The scope includes many different goals and concentration zones. Its main objective is to develop a sophisticated conversational AI system that is powered by the most recent large language model which is the GPT-4. The project looks at novel theatrical and entertainment applications, with the goal of producing AI actors or characters for original theatrical and entertainment experiences. The system is put to the test with actual users, and its performance is evaluated by looking at important criteria like accuracy, reaction time, and user happiness. The project also intends to share knowledge and ideas regarding the creation of the conversational engine with the academic and scientific community. Additionally, it considers

the potential ramifications for society and the broader societal impact of conversational engines driven by AI.

1.6 Report Structure

The report's content would be divided into several distinct chapters. Every chapter will focus on the project is in various stages, which are detailed below:

• Chapter 1 will discuss the project's fundamental structure, potential issues that might arise, and the project's main goals and objectives.

• Chapter 2 will discuss the many methods and suggestions that can be used to achieve this project's goals. There will also be a review of earlier, relevant research that was done by others.

• Chapter 3 will cover the methodology and the framework applied all through the project's development phase.

• Chapter 4 will cover the Development and design of the AI Powered Thespian and the timeline is showed in the terms of reference in appendix A

• Chapter 5 will provide the discussion of Implementation and Result.

•Chapter 6 will provide a conclusion to the overall report

Chapter Two - Literature Review

2.1 Overview

In the twenty-first century, artificial intelligence (AI) has expanded significantly and is becoming increasingly institutionalised. The definition of Artificial intelligence has come under close check in this period of interdisciplinary science, which includes automation, computer science, cybersecurity, mathematical logic, and linguistics. Researchers in the fields of computer science, mathematics, and engineering, began investigating the possibilities of artificial brains and attempting to characterise the intelligence of the machine as early as the 1940s and 1950s. Turing introduced the renowned "Turing Test" in 1950, which established the definition of "Machine Intelligence." Considering this, the workshop held on the campus of Dartmouth College in 1965 is where the roots of artificial intelligence can be found. McCarthy won participants over to his belief in the idea of "artificial intelligence" in this discussion. Additionally, it marks the start of the first "Golden age" of AI. With the ultimate objective of realising a world where humans and machines dwell peacefully, artificial intelligence (AI) strives to extend and enhance the capability and efficiency of mankind in activities including changing nature and regulating society through intelligent machines. (Xue et al., 2022) Since the 1980s, AI has been used in a few important fields, including natural language processing, computer vision, robotics, game theory, the science of cognition, machine learning and reason. This is due to the historical growth of AI. These topics evolved separately from one another. However, these fields had given up on the heuristic search-based techniques and logical reasoning that had been presented thirty years before. Rather, most of them were based on statistical methods. (J. Liu et al, 2018)

2.2 Conversational AI Agents

The study of artificial intelligence involves teaching computers to mimic intelligent human conduct by using them to imitate human learning, judgement, and decision-making processes. The goal of artificial intelligence (AI) is to emulate human intellectual activity by using knowledge as its purpose, knowledge acquisition, examination, and research of knowledge's ways of expression. In applications including speech recognition, image processing, natural language processing, the automatic demonstration of theorems, and intelligent robotics, AI has achieved outstanding accomplishments. AI combines several

different academic fields, including biology, psychology, philosophy, computer science, and logic, (Zhang and Lu, 2021).

The study of conversational AI has been ongoing for many years. Organisations from both academia and industry have a strong interest in these kinds of systems. The conversational AI system has significant commercial value and raises several intriguing issues, including those related to knowledge base reasoning, speech recognition, natural language processing, and designing human-computer interactions, among other things. Many large-scale conversational AI systems, including Siri, Xia ice, Alexa, and Google Assistant, have been developed.((Xue et al., 2022) Software agents known as conversational agents (CAs) enable human-computer interaction through natural language conversation. they are supported by an advanced engine that incorporates dialogue management techniques and natural language processing (O'Shea, Bandar, & Crockett, 2011).

Guidelines for ethical AI vary among organizations, illustrating the complexity of formulating guidelines for AI systems with societal impact. Disadvantaged users may experience disproportionate harm due to bias in various conversational AI applications. Critical analysis of social impact in publicly deployed conversational AI systems is hindered by the absence of cross-collaboration and the publication of proprietary training datasets and system architectures. In designing conversational AI,

Ruane et al. suggested that it crucial to consider linguistic elements, social and contextual factors, and the potential encoding of societal values and assumptions. Language and conversation nuances, often overlooked, should be regarded with equal importance as technical challenges in conversational AI development. Design choices in conversational AI development reflect specific worldviews and can influence the perception and handling of socially sensitive issues. While conversational AI in mental health services may seem appealing, it can carry potential harm due to limited capacity and inadequate ongoing evaluation of risks and benefits. Although ethical considerations are often mentioned in the evaluation of conversational AI, they are not consistently addressed explicitly. Addressing ethical concerns in conversational AI can be approached through safety engineering or machine ethics, but it necessitates a shift in mindset to consider the social context. Therefore, designers and developers must cultivate awareness of ethical issues and their impact on end users. This summary focuses on specific concerns for discussion and the clarification of implicit assumptions in conversational AI but does not cover all related concerns. (Ruane et al, 2019)

The Conversational AI architecture is composed of three (3) main components, each of which is further divided into more basic sections that conduct more basic functions. The first section deals with understanding user inputs made in natural language. The two NLU activities that comprise this activity are Entity Extraction and Intent classification. Intent classification aids the agent in comprehending the motivation behind the input. (Qiu et al, 2018). Request, inform, and place orders are a few examples of chatbot intentions in the food ordering space. In the healthcare space, similar examples include reporting a diagnosis, reporting a symptom, and requesting a prescription for medication. (J. Chen et al, 2018). The What of the input is dealt with through entity extraction. (X. Dong et al., 2016).

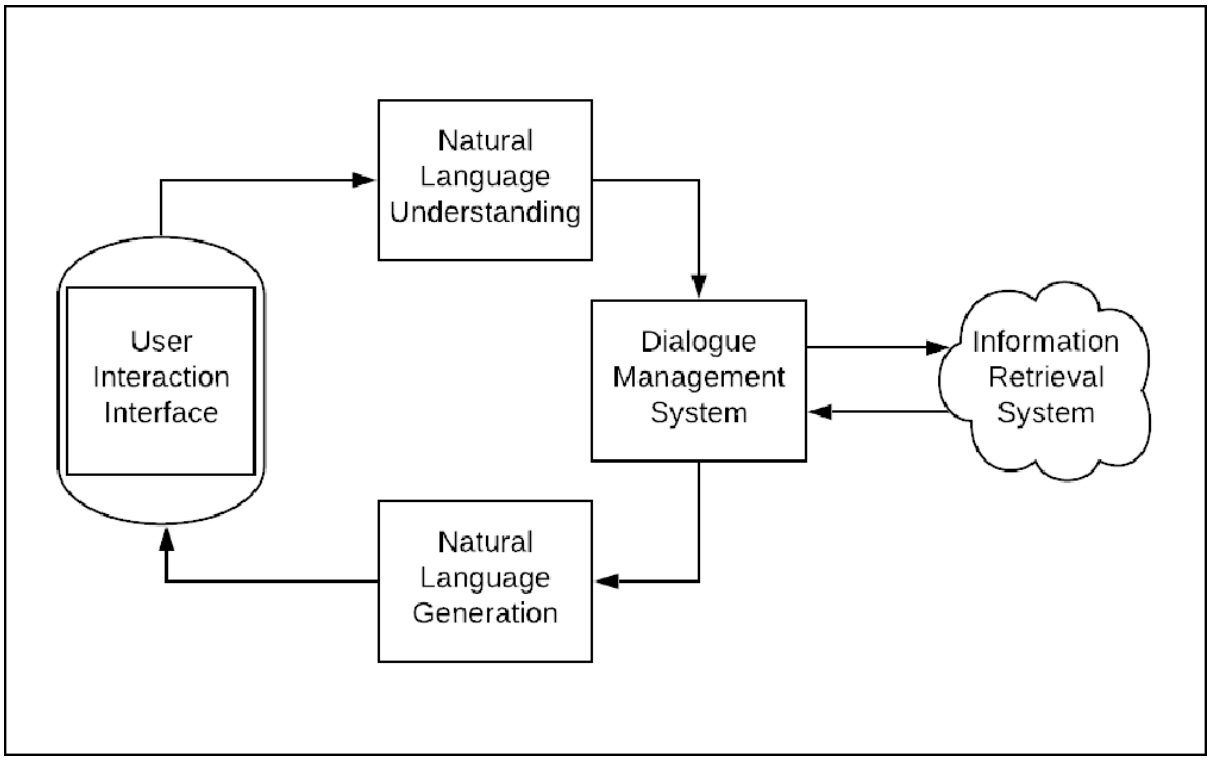


Figure 1.1: Conversational Engine Architecture

2.2.1 Natural Language Understanding

NLU enables the agent to identify the distinct informational units supplied by the user, which when combined with the user's purpose allows the agent to understand and comprehend the user's input. Thanks to newer, more effective methods of encoding natural language in a computer-understandable format while simultaneously defining relations between various entities, such as Word Embeddings, NLU has increased in accuracy over

some of its previous equivalents. Word embeddings is a word representation used in text analysis to take the form of a real-number vector, which captures the word's meaning in a manner that positions words with similar meanings closer together within the vector space. (J. Devlin et al., 2018). To correctly continue the discussion and avoid being stuck without the information it needs or in an incomplete circumstance, the Conversational Agent must decide on its own course of action after understanding the user's input. (Kulkarni, P et al, (2019).

2.1.2 Natural Language Generation

NLG concentrates on techniques for generating natural voice and language responses. In the context of conversational AI, natural language creation is essential for making the interaction seem more approachable and natural for the human user, which is a major element when evaluating the efficacy of conversational agents. According to Albert & Emiel's discussion (Albert. G. & Emiel. K., 2018), the Management system delivers data to the NLG module in a structured fashion depending on the history of the conversation and the present context. Therefore, the NLG component output is an NLP text or sentence, which also acts as the final output of the Conversational AI framework for each dialogue occurrence. The Natural Language Understanding and Dialogue Management Systems' processing and results form the basis for the NLG component's output. The promising techniques and the state-of-the-art techniques for NLG are as follows:

I. Sequence-To-Sequence (Seq2Seq):

Seq2Seq models have grown in prominence recently and are advertised as offering innovative performance in a range of tasks, including Conversational engines. The Seq2Seq models sequentially forecast tokens based on the previously acquired representation after translating a vector representation from an input sequence using LSTM models. The model uses a SoftMax function in equation (1.1) to successively predict tokens given inputs (X) and defines a distribution over outputs (Y). However, these Seq2Seq models have two common issues: between the train and test runs, there is measurement bias and inconsistency. (J. Li et al, 2015).

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^{k} e^{z_j}} \; for\; i = 1, \; . \; . \; , K \; and \; z = (z_1, \; . \; . \; , z_k) \in \mathbb{R}^K \qquad (1.1)$$

The standard unit SoftMax function $\sigma : \mathbb{R}^K \to (0,1)^K$ where $K \geq 1$ is defined by the formula in equation 1.1 (Wikipedia Contributors, 2019)

II. Sequence-To-Sequence and Reinforcement Learning

Yaser Kenshoo et al created a framework model for text summarization that combines the strength of reinforcement learning techniques with sequence-to-sequence models in natural language creation. They noticed that the reinforcement learning model needed only a short amount of time for training and that their work had an extremely high ROUGE score. Actor-Critic, Self-Critic, Policy Generator, E2E, and Argmax were the suggested models employed in the experimentation. The ROUGE (1, 2, L) scores for the Actor-Critic (Q-Learning) approach were (40.88,17.80,38.54), which is significantly higher than the average score for the other methods of (35.09,38.43,16.64). In a Seq2Seq and Reinforcement Learning method, the Seq2Seq model generates a collection of potential results that serve as the RL agent's action space, and the conversation history and context serve as the agent's state space or environment. (Y, Keneshloo et al, 2018).

2. 1. 3 Dialogue management

There has been lot of scholarly interest in a key element of the Conversational AI architecture that governs how the CA behaves and connects user input to the appropriate outcomes. The DM system oversees creating an interaction strategy that will direct the agent's decision-making process as it decides how to respond to user inputs. Research has received lot of attention over the past 20 years. Starting a conversation is the job of the dialogue management system. a plan that will help the agent make judgements on its own. based on previously submitted user contributions. There are two distinct kinds of DM systems: task-oriented systems and non-task-oriented systems. Task-oriented DM systems are responsible for moving the user from one state of the discussion to another to successfully fulfil a defined or dynamically understood task. Common examples of task-oriented conversational agents used today include those developed for straightforward tasks like booking a movie ticket, setting up meetings, managing projects, and answering FAQs. Task-oriented conversational bots can also perform more complicated tasks including medical diagnosis, visual question answering, and open domain question responding. Non-Task Oriented Systems are those created for casual,

unstructured, and human-like user interaction. Examples include chatbots that offer friendship rather than conducting jobs or transactions. (Kulkarni, P et al, 2019).

2.3 Ethical Considerations Towards Conversational AI

1. Privacy: The interaction between people and CAs, as well as the sporadic use of virtual agents like always-on appliances in homes, creates several ethical and legal concerns regarding the information that is acquired, who has access to it, how long it is kept, and where and how it is used. A variety of privacy issues are raised by the collecting of user data, some of which have a legal foundation and are handled by regionally specific data protection legislation, such the GDPR (General Data Protection Regulation) in Europe. The nature of these moral quandaries varies significantly depending on the domain in which the agent is deployed and the degree of receptivity of the user group. As AI systems are employed to produce decisions that are more significant, extensive, and are deployed in more areas of society, user privacy is essential and becoming much more essential. As a result, privacy should be viewed as a social issue that affects the entire society rather than just the individual user. Therefore, when selecting how to think about and regulate privacy, it is critical to understand how society may be impacted by the development of conversational AI systems. With the support of this viewpoint, privacy can be rethought in a way that connects it to aims and challenges that are more general to society. (Ruane, E. et al, 2019).
2. Transparency and Trust: Giving users control over how they choose to communicate with an agent is an essential first step in prioritizing their wants and wellbeing. It is critical to be transparent about an agent's status as automatic (non-human) and the limits of its capabilities, for example, to empower users to make informed choices and increase user confidence. Users may be better able to take charge of their own actions, particularly about information disclosure, if they are aware of whether they are conversing to an automated or human agent. This is especially important when sensitive user information, such that related to minors' education or banking, is being discussed or exposed, or when the conversation's implications and/or repercussions could have serious effects, including user health problems. Understanding user expectations of an agent is essential to preventing abuse of user confidence. It can be challenging to assess a system's behaviour in a way that allows us to verify whether the agent recommends goods based on true interests or requirements rather than unfairly categorising consumers based on factors like gender, race, age, or region. However, given the

potential degree of harm, it is crucial to constantly monitor and guarantee that users are not being profiled based on these delicate characteristics. In contrast to purchases of apparel or home goods, CAs that interact with users in a higher-risk situation, such as mental health services, have a greater social duty to their users and how the service may affect them. The user should always have assurance that the system will not exploit them and will act honestly when performing the advertised service. In order to achieve this, the agent must be evaluated to see how it is handling different sorts of users, and This calls for three things: (1) A thorough explanation of the agent's intentions and behaviours that the target user group can understand; (2) An evaluation to see how the agent is handling various user types; and (3) a grasp of the user's worries, expectations, and experience. (Ruane, E. et al, 2019).

3. Agent Persona: The highlighted passage underscores the significance of agent persona and personality in the decision-making process for Conversational AI (CA) design. These agent persona expressions encompass factors like gender, age, race, cultural background, and socioeconomic class, all of which can inform specific dialogue choices. The extent to which the agent embodies these characteristics can make them more or less evident in interactions. To determine whether the design of the agent's persona and related dialogues encourage potentially harmful actions, it is imperative to consider how the agent's persona may affect the kind of relationships users may wish to form with the agent. The design of agent personas can unintentionally perpetuate detrimental stereotypes, especially considering that many publicly available agents tend to present as female, which may align with market research but can reinforce socially ingrained gender biases. This gendering of CAs may result in female personas being used in submissive contexts while male personas are utilized in authoritative situations. In certain domains, there is a growing trend toward androgynous agent personas, as exemplified by the banking agent Kai. (Ruane, E. et al, 2019).
4. Anthropomorphism and Sexualization: Machines are frequently anthropomorphized by humans. When users may have verbal conversations with a system, especially if it has been given personality and embodied through an avatar or another method, this type of anthropomorphism is accentuated. This has been observed throughout history, even when the creators themselves are against anthropomorphizing and exaggerating the capabilities of technology. For instance, ELIZA's founder Joseph Weizenbaum (1964-66) said unequivocally that ELIZA was unable to communicate with full understanding. Nevertheless, many users were persuaded by ELIZA's wit and compassion. Unwanted

romantic interest in the agent may be a surprise facet of human-computer interaction. The well-known entertainment chatbot Mitsuku1, which has won the Loebner Prize four times, is a nice illustration of this. The developer and maintainer of Mitsuku, Steve Worswick, has talked about the kind of romantic attention "she" receives and even the letters he gets from users pleading for her release. (Ruane, E. et al, 2019)

2.4 Role and Integration of conversational AI in Theatre

Annabel Latham explained that conversational engines are supported by a cutting-edge conversational agent that employs natural language processing and dialogue management methods to preserve context throughout the conversation and to enable initiative interaction (i.e., Users can ask queries without following a predetermined sequence.) (Jackson, D., & Latham, A. 2018).

Interactive fiction systems have been the focus of conversational agent (CA) development in the storytelling domain. At first, natural language processing models were used by early interactive fictions like Adventure to enable players to engage with the plot by responding to straightforward text commands like "go north" or "take apple." To allow users to converse, agent-based storytelling experiences like Lyotard and Galetea implemented restricted fluency CAs using IF authoring frameworks. However, these systems relied on complex rules to simulate character responses, resulting in poor interactivity. Users often struggled to find specific word combinations that would prompt an output from the agent to progress the story. The primary application of CAs in storytelling has been within computer games, where players interact with non-playable characters (NPCs). Typically, these interactions are limited to multiple-choice dialogue options that affect gameplay decisions. The goals and challenges of these games extrinsically encourage the conversational style. As a result, if NPC interactions do not increase players' chances of winning the game, they are unlikely to appreciate them. Commercial solutions that enable the development of Storytelling CAs (SCAs) as a component of digital media work have lately come into existence. Systems like Charisma.ai were created to make it possible for drama-based applications to "break the fourth wall" and communicate with users directly. These systems also incorporate extra AI features like emotion sensing to improve emotional connection to the story's characters. SCA usage has increased in areas involving cultural heritage as well for instance, the National Holocaust Centre in the UK offers people the chance to interact with digital recordings of Jewish Holocaust survivors. Using a microphone, users can ask the SCA questions, and the

system's huge database of 800 to 1,200 questions and answers enables it to provide prompt and insightful answers. The Forever Project serves as an example of how SCA applications might help students and people or characters in historical and cultural contexts establish a natural and engaging relationship. (Kulkarni et al, 2020)

Chapter Three - Theoretical Framework

3.1 Natural Language Processing

The ability of computers to recognise and understand human text language is known as natural language processing. AI's goal is to process natural language, and the biggest difference between natural language and artificial language is that the former is founded on human thought. The study of making computers "understand" natural language is known as computational linguistics, which is situated at the intersection of language information processing and artificial intelligence. The secret to understanding natural language is this. It starts out as a machine gathering sound signals. The meaning of the text is translated from sound signals to text signals using natural language processing technologies. Then, a machine converts sounds into words and words into meanings. Once these two phases are complete, the machine can hear and comprehend. The algorithm in continuous learning is optimised by the machine's speech recognition and semantic understanding technology so that it cannot only listen but also understand - and even grasp emotions. (Zhang & Lu, 2021).

Users who do not have the time to learn new languages or to hone their current ones can benefit from natural language processing (NLP). The objective of NLP, a subfield of linguistics and AI, is to make it possible for computers to understand claims and words used in human languages. It was created to simplify the user's task and to allow them to communicate with a computer in natural language. The two categories that advance the process of comprehending and generating the text are natural language generation and natural language understanding. Phonology, which studies sound, morphology, which studies word formation, syntax, which studies sentence structure, semantics, which studies syntax, and pragmatics, which studies understanding, are all included in the study of language Noah Chomsky, one of the first linguists of the twelfth century to establish syntactic theories, held a distinctive place in the realm of theoretical linguistics since he revolutionized the study of syntax (Chomsky, 1965). Natural language generation, or NLG, is the act of constructing meaningful words, phrases, and paragraphs from an internal representation. (Khurana, D. 2023)

The rationalist or symbolic approach assumes that a massive portion of knowledge in the human mind is fixed in advance, through genetic inheritance, rather than being acquired through the senses. The person who supported this strategy the most was Noam Chomsky. It was once thought that by providing some basic knowledge and reasoning capabilities, robots might be made to work similarly to the human brain. For example, it was thought that

knowledge of languages is directly embodied in rules or other kinds of representation. This facilitates natural language processing on an automatic basis. Algorithm evolution is a key component of statistical and machine learning, which enables a programme to detect patterns. The fundamental algorithm of a given method, which is optimised by a numerical measure that characterises numerical parameters and learning phase, is characterised by an iterative process. The two main types of machine learning models are generative and discriminative methods. Generative methods focus on modelling probability distributions and can be used for generating artificial data. They are known for their ability to produce detailed models of data, including the underlying probability distribution. Discriminative methods, on the other hand, are more practical and are primarily concerned with distinguishing between different classes or categories based on observed features. Each type of model has its unique characteristics and use cases. Generative models can, for example, be utilized for identifying the language of an unknown speaker without extensive prior language knowledge, as they model the underlying data distribution. However, they may face challenges when dealing with many characteristics or features. Discriminative techniques, such as logistic regression and conditional random fields (CRFs), are better suited for tasks like language differentiation. They focus on identifying patterns that distinguish one class from another and are less knowledge-intensive than generative models. Examples of generative methods include Naive Bayes classifiers and hidden Markov models (HMMs), while logistic regression and conditional random fields (CRFs) are examples of discriminative methods." (Khurana et al., 2023)

3.2 Foundation Models

The most significant advancement in foundation models has occurred in NLP. We view foundation models as a universal paradigm of AI rather than being in any way specific to NLP, like how deep learning, which gained prominence in computer vision but applies to other fields, is viewed by many as a paradigm of AI. The NLP community was poised for another seismic upheaval by year's end in 2018 that would usher in the era of foundation models. Technically speaking, foundation models as shown in figure 3.1 are made possible by scale and transfer learning. Transferring "knowledge" from one task—like object detection in photos—to another—like activity recognition in movies—is referred to as transfer learning. Pretraining, which involves training a model on a different task (often just till the finish) and then fine-tuning it to adapt to the desired downstream job, is the most widely used technique for transferring learning. Although scale is what gives foundation models their power, transfer learning makes them practical. The scale needed the next three components. The architecture

of the Transformer model (Vaswani et al. 2017) was created to take use of the parallelism of the hardware and the availability of more training data to train models that are much more expressive than before. For instance, over the past four years, GPU throughput and memory have improved by a factor of one. the importance of having access to and using data. In contrast, the pretraining job in self-supervised learning is automatically created from unannotated data. For example, in the masked language modelling task that was used to train BERT (Devlin et al. 2019), participants were asked to guess the word that was missing from a sentence based on the context of the rest of the sentence (for instance, I like sprouts). Self-supervised tasks force the model to anticipate parts of the inputs, which makes it more complicated and potentially more useful than models trained on a narrower label space. As a result, they are not only more scalable because they simply need unlabelled data, but they also make the model more complex. Modern NLP models are now mostly descended from a few numbers of foundation models, including as BERT, BART, etc., which have led to a level of standardisation never seen. Although this homogenization has an incredibly high leverage (any modifications in the foundation models can immediately help all of NLP), a drawback is that all AI systems may acquire the same problematic biases from a small number of foundation models. Additionally, there is a trend towards uniformity throughout research communities. An unanticipated development that comes from size has also been facilitated by foundation models. For instance, GPT-3.5 (Brown et al. 2020) allows for in-context learning, where the language model can be adapted to a downstream task by giving it a prompt (a natural language description of the task), an emergent property that was neither specifically trained for nor anticipated to arise. GPT-3.5 has 175 billion parameters compared to GPT-2's 1.5 billion. (Bommasani et al, 2021)

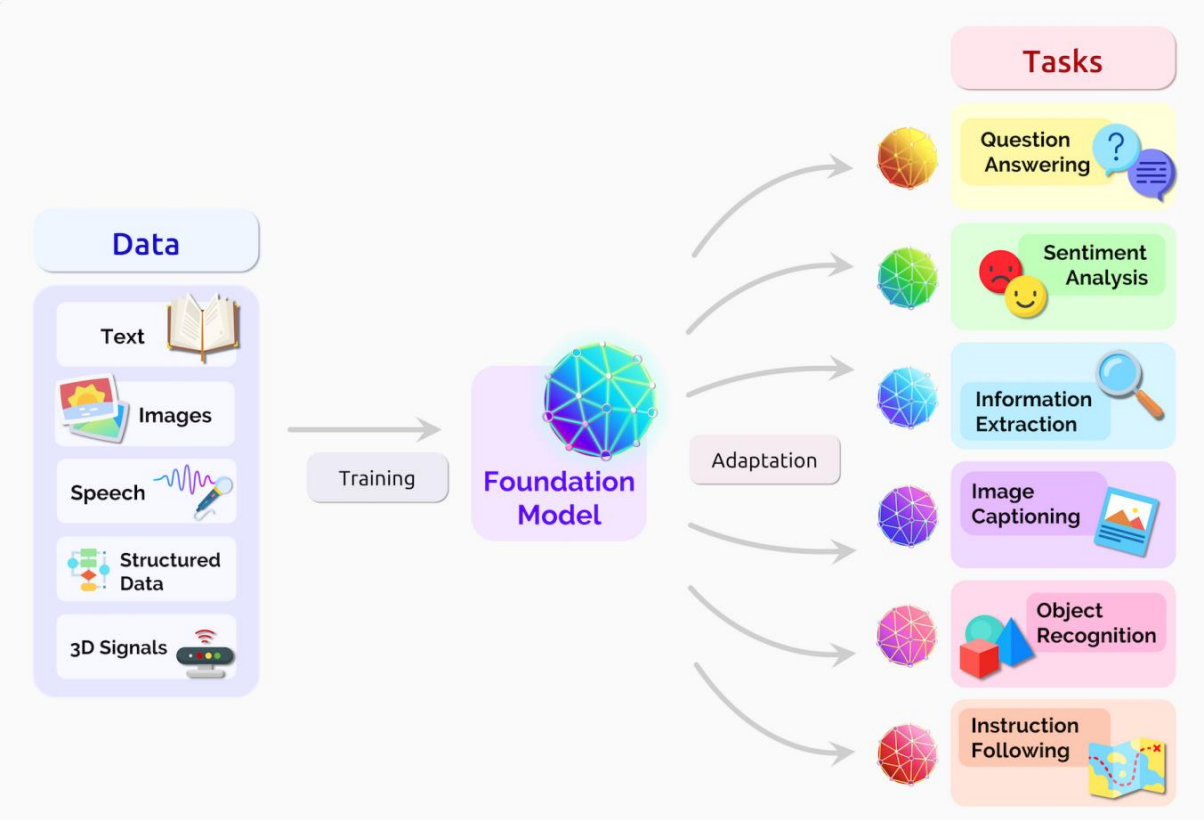


Figure 3.1: foundation models showing all the data from various modalities.

3.3 Large Language Models

Vast language models (LLMs) often refer to Transformer language models, such as GPT-3, GPT-Neo, GPT-NeoX, GPT-J, PaLM, Galactica, and LLaMA, which are trained on vast amounts of text data and have hundreds of billions (or more) of parameters, which has given new meaning to the phrase "unreasonable" effectiveness of data". Figure 3.2 shows the recently developed large language models. (Zhao et al, 2023). The effectiveness of LLMs is "unreasonable" in three interrelated ways:

• First, there are qualitative leaps in capacity as the model's scale, and

• Second, the performance of LLMs on benchmarks scales with the size of the training set and model size.

• Third, with a powerful model, many tasks that need intellect from people can be simplified to next token prediction.

Large language models (LLMs) are generative mathematical models that analyse the statistical distribution of tokens in human-generated text, and they can be used to predict the words to follow a given sequence of text based on the statistical distribution of words in the public corpus. LLMs find application in conversational agents or AI assistants like ChatGPT,

highlighting the importance of distinguishing between LLMs themselves and the systems they are embedded in for philosophical clarity. It is important to note that LLMs generate statistically likely sequences of words and do not possess knowledge, understanding, or consciousness. When evaluating the utility of LLMs, it is essential to consider the distinction between real-world facts and fictional information. Therefore, users and developers should be aware of what LLMs really do to avoid misleading descriptions of their capabilities. While LLMs can provide accurate responses based on statistical patterns in the public corpus, their answers may not always align with real-world truth. Hence, understanding the limitations and functions of LLMs is crucial when using or developing them. As LLMs become more capable of mimicking human language, it becomes increasingly tempting to anthropomorphise them (Shanahan, 2023)

We highlight the LLMs in yellow using publicly accessible model checkpoints. We only include the LLMs with review outcomes that have been made publicly available due to the space restrictions of the figure.

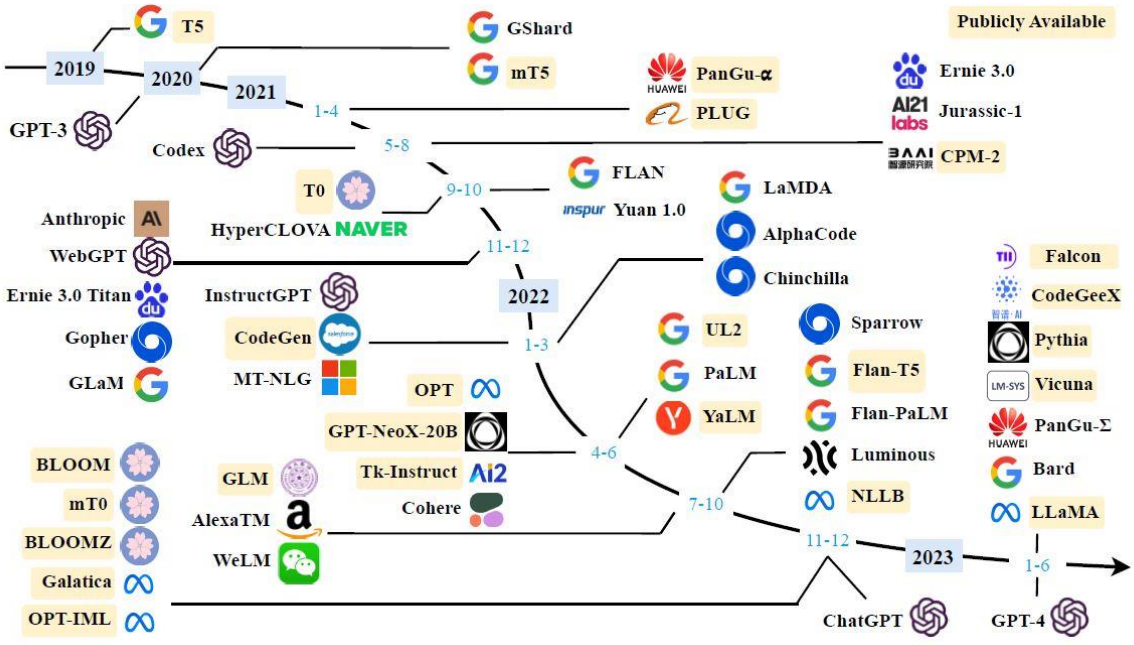


Figure 3.2: a timeline of recently developed big language models (greater than 10B in size).

In the context of large language models and machine learning there are learning approaches used to mitigate the challenges of training machine learning models when there is limited labelled data for a specific task. They include:

- Few-shot learning describes a situation in which the model receives a limited number of task demonstrations at the time of inference, but no weight modifications are permitted. Few-shot learning's key benefits are a decreased demand for task-specific data and a decreased risk of learning an excessively narrow distribution from a large but limited fine-tuning dataset. (Brown et al, 2020)
- One-shot learning only allows for one task demonstration and a natural language description as seen in figure 3.3, one-shot learning only allows for one task demonstration and a natural language description. (Brown et al, 2020)
- Zero shot learning just requires a description of the job in natural language rather than any demonstrations. The most demanding situation is zero shot learning, which may be "unfairly hard" since it might be challenging for people to comprehend the structure of the job without prior instances. (Brown et al, 2020)

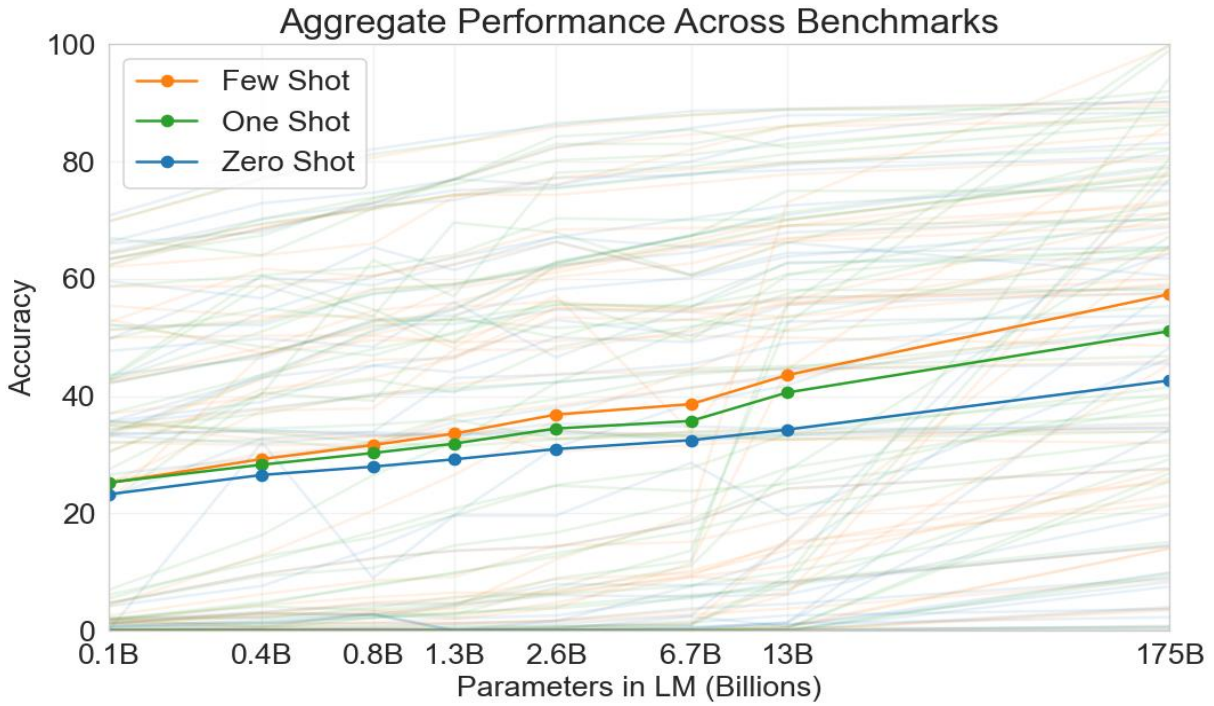


Figure 3.3 – Zero shot, one shot and few shot learning on LLMs.

3.2.1 Machine Learning and Deep Learning Processes

The study of machine learning focuses on the creation of statistical models and algorithms that let computers conduct activities without having to be explicitly programmed. Many of the programs we use every day, like web search engines like Google that utilize a learning algorithm to rank web pages, use learning algorithms. Data mining, image processing, predictive analytics, and other processes all involve machine learning techniques. Most AI systems today are powered by machine learning, which develops predictive models using

previous data and then applies those models to make predictions about the future. Machine learning, which first appeared in the 1990s and marked a substantial change in how AI systems were formed, contrasted with the way AI systems were previously built. A learning algorithm would infer how to perform a task based on data rather than providing instructions, so the how emerges from the dynamics of learning. Additionally, machine learning Because they allowed the use of a single generic learning technique, such as logistic regression, to power a variety of models, we referred to these models as "foundation models" to describe their unfinished but vital status. (Bommasani et al., 2021).

3.2.2 Fine Tuning and Instruction Tuning

Fine-tuning is a method of training a pre-existing model on a specific task by using a supervised dataset that is specific to that task. This method requires thousands to hundreds of thousands of labelled examples to achieve impressive performance on benchmarks. While fine-tuning has the advantage of achieving powerful performance on many benchmarks, it also has a disadvantage which is the potential for poor generalization out-of-distribution, which means that the model may not perform well on data that is different from the training data. Fine-tuning can also exploit spurious features of the training data, potentially resulting in an unfair comparison with human performance. (Brown et al, 2020)

Instruction tuning, which is closely related to fine-tuning and multi-task prompted training, is the process of optimising pre-trained LLMs on a set of structured examples in the form of natural language. We must first gather or create instruction-formatted instances to do instruction tuning. Using these structured instances, we next use supervised learning techniques (such training with the sequence-to-sequence loss) to improve LLMs. Even in a bilingual environment, LLMs can show higher generalisation skills after instruction tailoring.

3.2.3 Prompt Engineering

Prompt engineering is a new profession that concentrates on developing and optimizing prompts, with the goal of effectively using language models (LMs) for a wide range of applications and research topics. It is simpler to understand the promise and limitations of large language models (LLMs) if one has quick technical skills. Prompt engineering is a strategy used by developers to create reliable and efficient prompting methods that interface with LLMs and other tools. Not merely creating prompts is what prompt engineering entails. It includes a wide range of abilities and methods that are helpful for communicating and working with

LLMs. The ability to communicate with, collaborate with, and comprehend LLM skills is crucial. (Saravia, 2022) Different ways to prompt a conversational agent or a LLM include:

- Think Step by Step

Simply adding "Think step by step" at the end of a prompt is one of the most well-known (and straightforward) prompt engineering approaches. The University of Tokyo and Google researchers discovered that including this sentence increased the accuracy of the GPT-3 (text-davinci-002 model) on several tests. On the MultiArith test, for instance, accuracy went from 17.7% to 78.7% (Kojima 2022). MultiArith problems are arithmetic problems that must be solved in several phases. (Gao, 2023) The effectiveness of think step by step was prompted using GPT 3.5 in figure 3.4 and gave the accurate output.

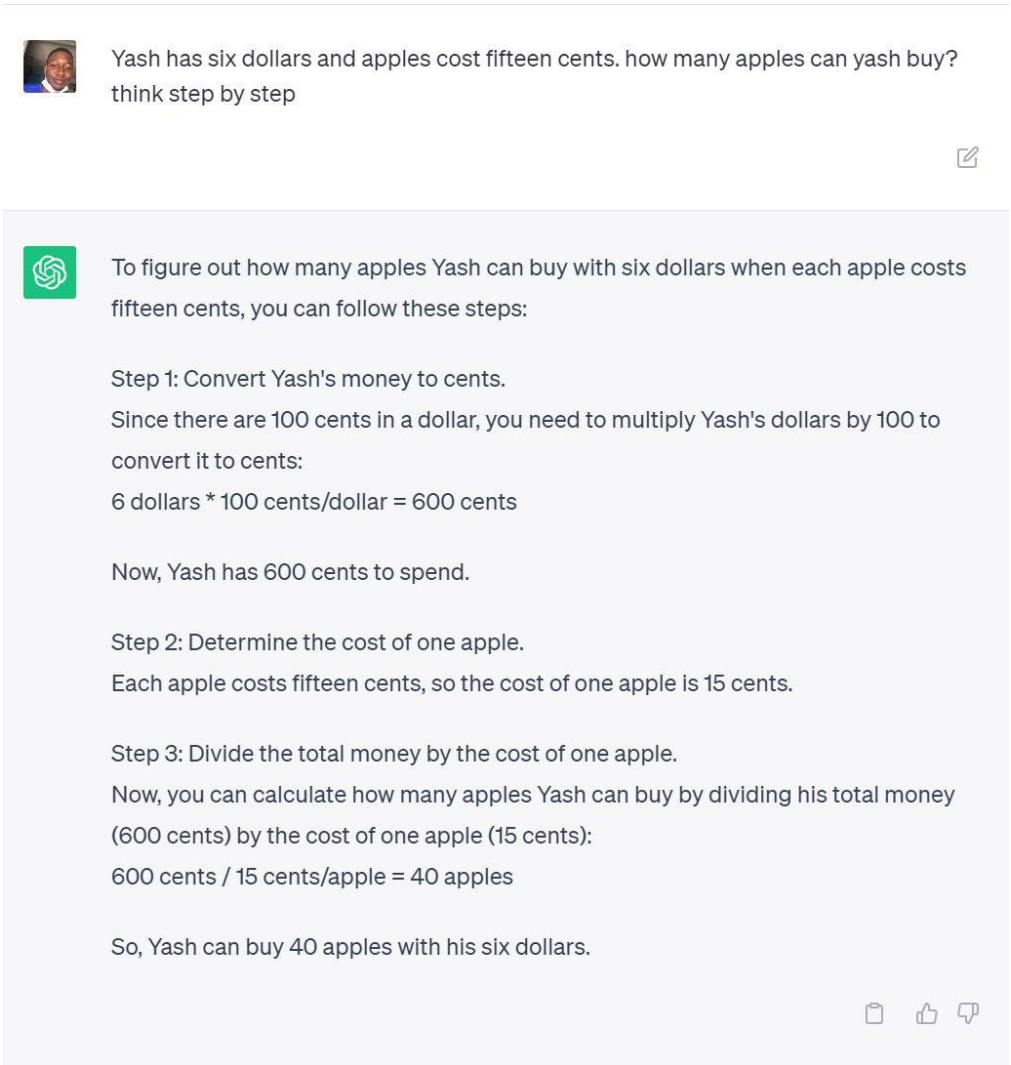


Figure 3.4 – Think step by step using GPT 3.5 Turbo

- Role Prompting

Role prompting is another tactic that can be used to encourage LLMs to produce material in a particular creative style, such as an author. Some users claim that when they let Large Language Models (LLMs) know that they are subject-matter experts, the outcomes improve just as the prompt given to ChatGPT in figure 3.5. When the user requests code, this is accomplished by prepending a prompt to the LLM, such as "You are a medical Assistant." As seen in figure 3.5. One explanation for this tactic is that it aids the LLM in concentrating and helps it decide which areas of its information to prioritise and place first. This is because, although having a great amount of knowledge, LLMs could not be aware of the portions that are pertinent to the user's prompt. Knowing the user's area of expertise will enable the LLM to produce more pertinent and precise responses. This is helpful when creating content for industries like medical, education, marketing, or art and entertainment where a particular tone or style is sought.

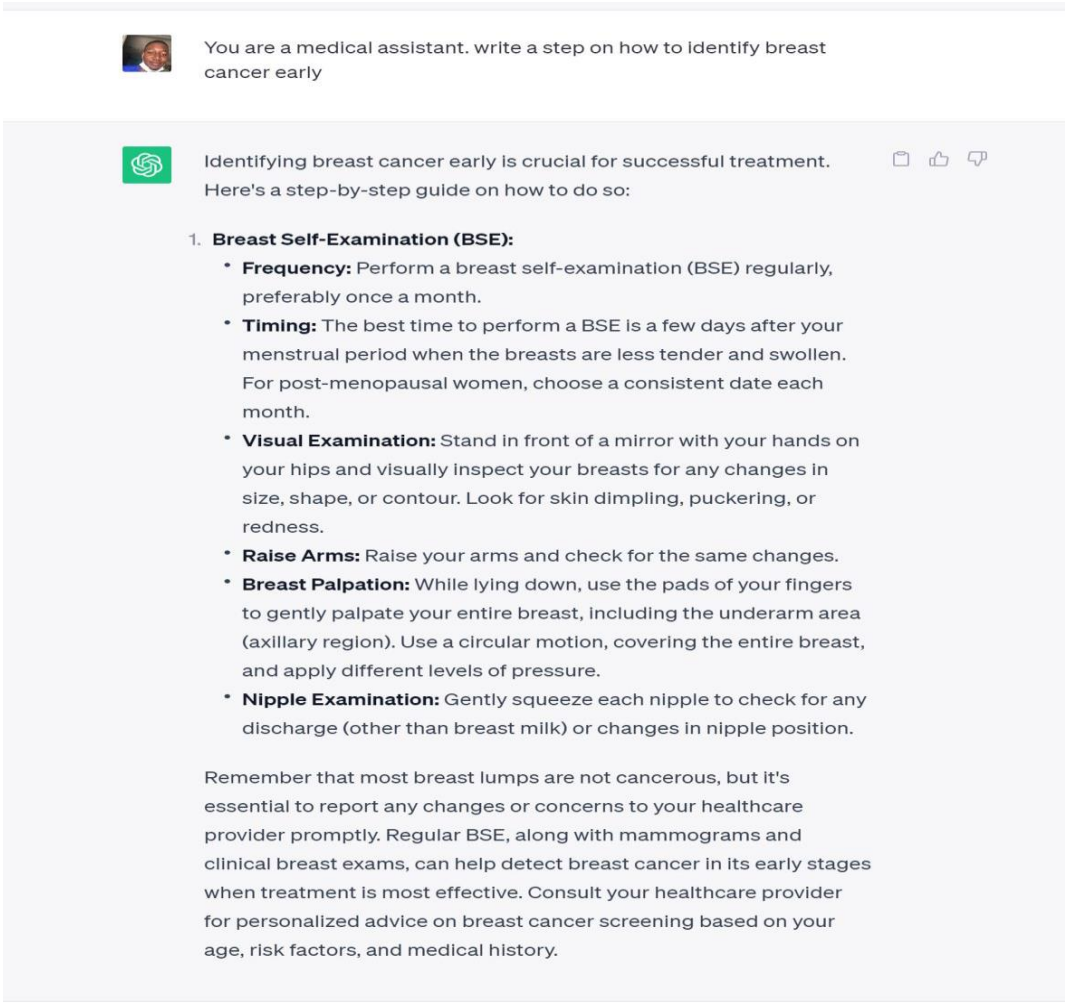


Figure 3.5 – Role Prompting using GPT 3.5 Turbo

- Few-shot Learning.

Few-shot learning involves providing examples of desired outputs to a language model, which helps it produce the desired outputs by constraining the potential output space. It is important to provide diverse and balanced examples to avoid bias and ensure coverage of all scenarios of interest. To avoid bias and ensure accuracy, it is important to provide diverse and balanced examples when training GPT for sentiment classification, covering all desired scenarios including neutral classification as seen in figure 3.6.

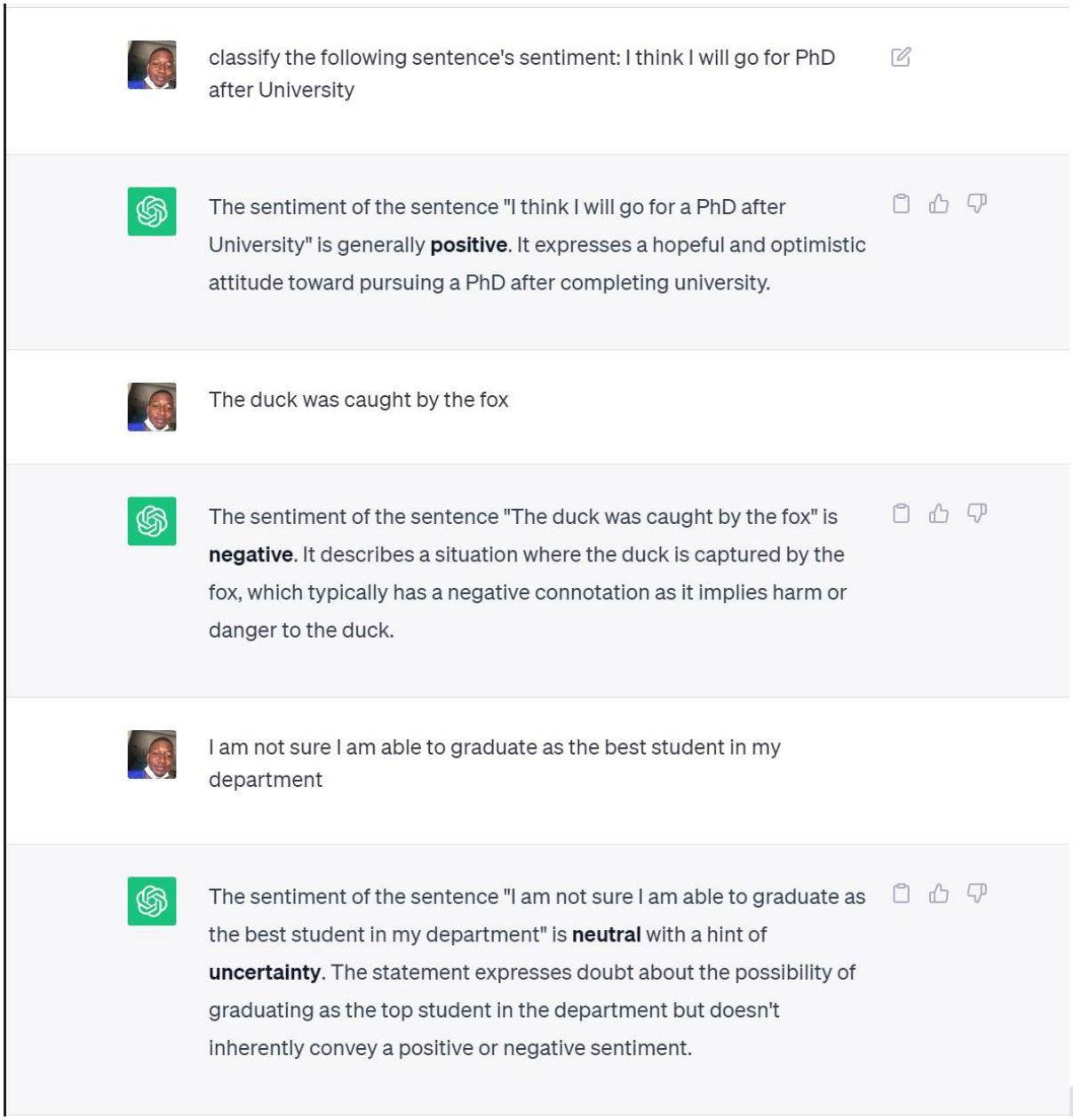


Figure 3.6 – Few shots learning through sentiment analysis using GPT 3.5 Turbo

3.4 Speech Recognition

Speech recognition is a technology that has emerged in the twenty-first century and has become an integral part of our lives. The application of speech recognition has made our lives more comfortable and effective by enabling us to interact with machines using our natural voice. Speech recognition is a technique that allows machines to identify and understand human speech signals. This means that machines can interpret and respond to human speech, creating a communication link between humans and machines. The process involves the use of an acoustic and modelling algorithm which helps the machine to analyse and interpret the speech signals it receives. as seen in Figure 3.7 shows the acoustic algorithm which is responsible for converting the speech signals into digital format that can be processed by the machine. The modelling algorithm then uses statistical models to analyse the speech signals and identify the words and phrases being spoken. Once the words and phrases have been identified, the machine can then respond appropriately, either by providing information or conducting a specific task. (Ibrahim & Varol, 2020)

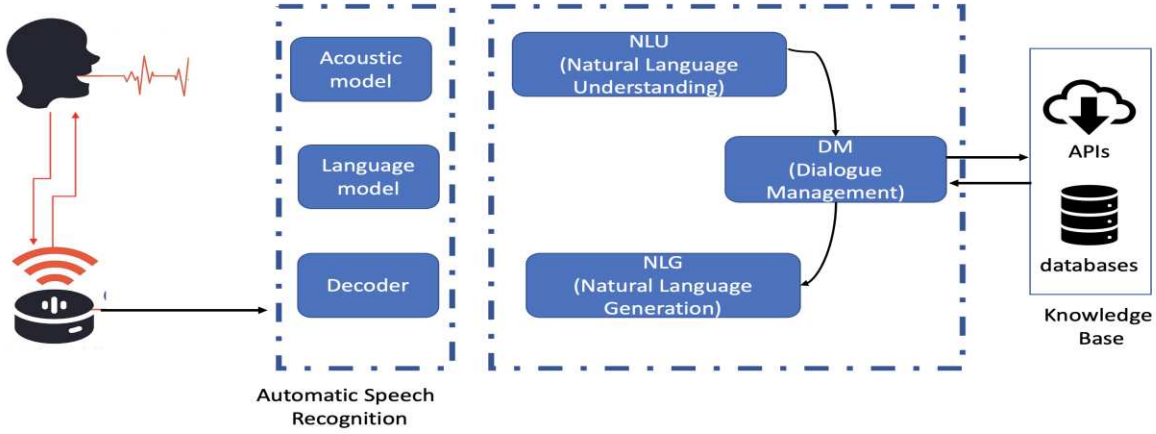


Figure 3.7: High level architecture for voice-based conversational AI agents

The automatic speech recognition system (ASR) is a technology that aims to create a system that can understand human language and translate speech into a different medium for easy interpretation and understanding. The ASR system uses an acoustic model, modelling algorithm (language model) and decoder to execute its task of identifying the human voice through speech signals. The ASR system can convert speech into text, making it easier to interpret and understand, especially for people with hearing impairments. The development of ASR technology has been a significant achievement in the field of computer science, enabling machines to interact with humans in a more natural and intuitive way. (Ibrahim & Varol, 2020)

Chapter Four - Design and Methodology

4.1 Problem Definition and Target Audience

4.1.1 Problem Definition

The issue at hand is the requirement for a novel conversational engine that makes use of the capabilities of AI, in particular GPT-style APIs, to enable imaginative and participatory storytelling. Traditional AI conversational systems can lack the nuance and originality necessary for compelling narratives and lively discussions. We want to create a software solution that not only engages users in on-the-spot chats but also expertly creates interactive stories to solve this. The specific challenges will want to overcome include:

- Speech-to-speech operation is difficult because it entails turning spoken language into text, processing that text, and producing intelligent speech responses.
- The conversational engine's performance should be able to store and analyse conversational interactions to be fine-tuned in the future, customised to meet the preferences of specific users, and improved overall.
- Unavailability of Data: the chatbot will leverage on the latest Large Language Models which is the Generative Pre-trained Transformers 4 having over one trillion parameters.

4.1.2 Target Audience

The target audience for this innovative conversational engine is diverse and comprises of individuals and professionals across various domains which include:

- Content Creators: writers, copywriters, and content creators can use this app in helping them to develop content ideas and refine their own content ideas into something much better.
- Actors: actors can use this app to practice their scenes, research on a character, prepare for auditions and offer constructive feedback around their roles.
- Film Directors: film directors can use this app to generate scripts, create monologues, analyse scenes, and give ideas around the film production.
- Educators: Teachers in the fields of digital art, filming, theatre interested in enhanced learning experiences through AI content and storytelling will find this app of significant use.
- General Users: Anyone who wants to converse with the app in a meaningful and enjoyable manner can use the app for companionship and read interesting stories they ask the app to share.

- Entertainment Industry: Game developers and entertainment businesses seeking AI driven interactive storytelling for games, virtual reality experiences, cartoons, animations will benefit from the app.

4.2 Programming Languages

Programming languages serve as means of communication with computers to instruct them to conduct specific tasks. These languages encompass numerous functionalities that computers can comprehend and execute. Within this development context, two primary languages employed are Python and JavaScript.

Python is a widely recognized general-purpose programming language celebrated for its simplicity and readability, rendering it a highly suitable choice for application development. In the context of this project, Python was leveraged extensively to manage most of the backend codebase.

JavaScript, on the other hand, stands as a versatile programming language particularly well-suited for web development. Its capabilities extend to crafting interactive and dynamic web pages, cementing its indispensability in frontend web development. Furthermore, TypeScript, an augmented subset of JavaScript, played a pivotal role in the frontend development phase. TypeScript's incorporation enhanced code maintainability and scalability, making it an instrumental addition to the development process.

4.3 Software Development Process

Software development is the ideation, specification, design, coding, documentation, testing, and bug-fixing processes involved in creating and maintaining applications, frameworks, or other software components. Software development, which typically involves a planned and methodical process and frequently overlaps with software engineering, broadly refers to all actions from the conception of the desired program to its eventual embodiment. The source code must be written and maintained. Software development also includes any other tasks that lead to the creation of software, such as prototype, modification, reuse, re-engineering, and maintenance. Figure 4.1 displays the architecture of the development which encompasses both frontend and backend development. The application currently under development is an AI-powered Conversational Engine available as both a web and mobile application. This application takes both speech and text inputs and provides responses in the form of speech and text. The Conversational AI app is designed to create interactive stories and serve as an assistant in various fields, including healthcare.

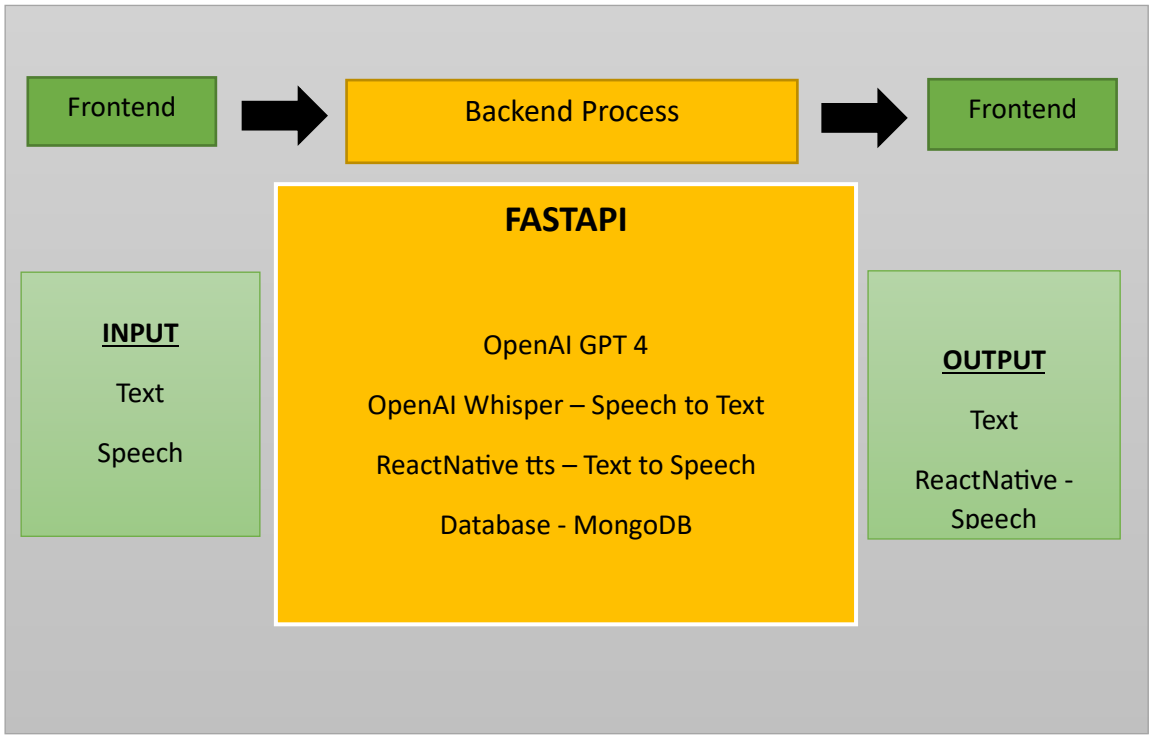


Figure 4.1: System Architecture for Software Development

4.4 Backend Development

The Backend development is the backbone of the entire process.it manages the application architecture and sever-side scripting. It is responsible for building and maintaining the back end of the application.

In this app, the backend focused on developing the databases, back-end logic, application programming interface (APIs), architecture and servers. They utilize code that facilitates communication between browsers and databases, enabling the storage, comprehension, and removal of data.

4.4.1 Application Programming Interface

Application programming interfaces (APIs) allow two or more computer programs to communicate with one another. It is a type of software interface that gives other software programs a service. In contrast to a user interface, which links a computer to a human, an application programming interface (API) connects computers or pieces of software to one another. It is solely intended for usage by the computer programmer who is integrating it into the software, not by the end user. The many parts that make up an API for a programming language usually function as tools or services for programmers. When a program or

programmer makes use of an API, that piece is referred to as being invoked. Calls also apply to the requests or endpoints that make up the API. A specification for an API defines these calls and explains how to utilize or implement them. (Wikipedia). The API used in this project is the latest OpenAI GPT 4. As seen in the code below, the function (text to text response) holds the OpenAI API which parses arguments like model (GPT4) that initiate the kind of model to use, the max tokens are set to two hundred (200) which is approximately about 175 words taken by the input. The temperature determines if the output will be random, it is used to adjust the probability distribution of the predicted output. Since we need to get more specific output and not just some random results, we use the lowest temperature which is set at 0.1. the messages take the system and the user which can be in this format.

Character = "You are a psychology Lecturer with a PhD in Psychology"

Prompt= "How can I master emotional intelligence"

```
def text_to_text_response(text_input):
    system = {"role": "system", "content": character}
    user = {"role":"user", "content": prompt}
    try:
        response = openai. ChatCompletion.create(model="gpt-4", max_tokens=200, temperature=0.1,
messages= [system, user])
        result = response["choices"][0] ["message"] ["content"]
        return result.
    except Exception as e:
        return ("Failed to get text response from GPT4 API")
```

4.4.2 OpenAI GPT

The Generative Pre-Trained Transformers (GPT) is a state-of-the-art large language model developed. The GPT-4 was used for this development. The numerical values known as parameters control how a neural network processes input data and generates output data. They are encoded with the model's knowledge and abilities and are learned from data during the training phase. A model can be more complicated and expressive as well as manage more data if it contains more parameters. (OpenGenus IQ,2023) The GPT 4 has about one trillion parameters, and it is the latest of the GPT Foundation models. Enhancing these models' capacity to comprehend and produce natural language content, particularly in increasingly intricate and nuanced contexts, is one of the key objectives of their development. GPT-4

outperforms existing large language models on a collection of NLP tasks, and exceeds the vast majority of reported state-of-the-art systems (which often include task-specific fine-tuning) (OpenAI, 2023)

4.4.3 Fast API

Fast API is an innovative web framework for creating Python RESTful APIs due to its simplicity, speed, and resilience. It was chosen for this development. IT is strong for production situations because it completely supports asynchronous programming and runs on Gunicorn and (Asynchronous Server Gateway Interface) ASGI servers like Uvicorn and hypercorn (FastAPI, 2023).

FastAPI made use of HTTP communication protocols which connected the backend to the frontend of the development such as GET (request information), POST (send information), PUT (update/append information), DELETE (remove information) which are used to create CRUD Operations. In this development, as seen in the code below the POST and GET methods were used to create route /text and /history for the frontend communication.

```
@app.post("/text") ←
async def post_text (email, textinput):
    message = text_to_text_response(textinput)
    return message.
@app.get("/history/{email}") ←
async def get_chat_history(email):
    return {"chat_history": chat_history}
```

You can interact with API endpoints right on your local machine without any external server. On running the command *"uvicorn main:app –reload"*

you should see a success message stating that the server has started as shown in figure 4.2 and the browser is fired up to take the URL http://127.0.0.1:8000/docs. The output that you will see will be the JSON response directly from the server which is displayed in figure 4.3. Appendix B4 shows the deployed version of the backend application.

```
(venv)
ogund@RichieDiamond MINGW64 ~/Downloads/thespain (main)
$ uvicorn main:app
INFO:     Started server process [19824]
INFO:     Waiting for application startup.
INFO:     Application startup complete.
INFO:     Uvicorn running on http://127.0.0.1:8000 (Press CTRL+C to quit)
```

Figure 4.2: Uvicorn on Terminal

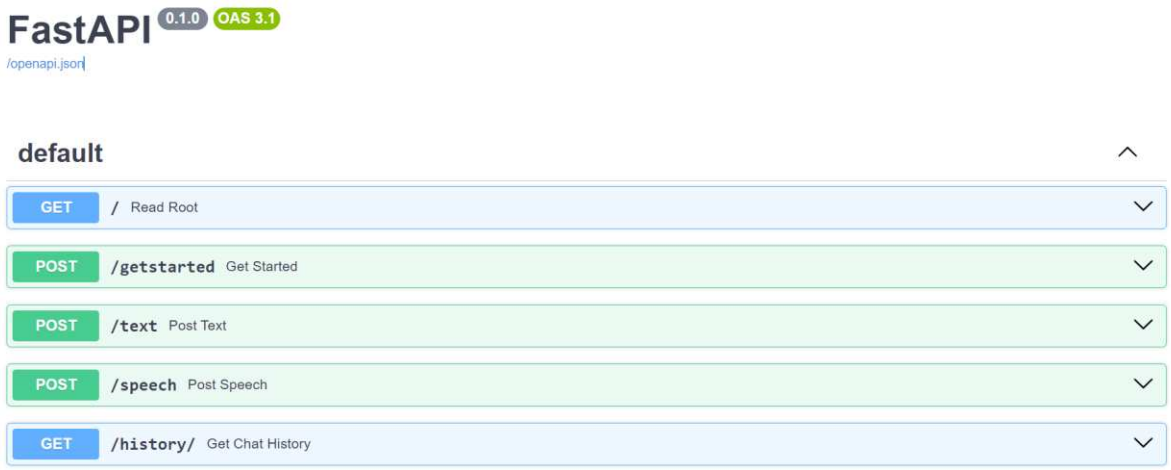


Figure 4.3: Uvicorn Interface

4.4.4 OpenAI Whisper

Whisper is an automated speech recognition (ASR) system that has been trained on 680,000 hours of supervised, multilingual, and multitask online data. We demonstrate how using a dataset this size and variety increases robustness against accents, background noise, and technical terminology. Additionally, it permits both translation into English from several languages as well as transcription in those languages. We are sharing our models and inference code to lay the groundwork for future research on robust speech processing as well as for the creation of useful applications. The encoder-decoder Transformer is how the Whisper architecture is implemented, which is a straightforward end-to-end strategy. An encoder receives input audio that has been divided into 30-second segments and transformed into a log-Mel spectrogram. With the help of specific tokens that instruct the single model to conduct tasks like language identification, phrase-level timestamping, multilingual voice transcription, and to-English speech translation, a decoder is trained to anticipate the matching text caption. (OpenAI, 2023)

Whisper played a pivotal role in this development project by serving as the primary speech-to-text conversion tool. Its advanced speech recognition capabilities allowed for the seamless transformation of spoken language into textual data. By harnessing the power of Whisper, the project was able to transcribe verbal input accurately and efficiently, thus enabling enhanced data processing, analysis, and accessibility within the system. This integration of Whisper not only streamlined data input but also significantly improved the overall user experience by providing a robust and reliable mechanism for converting spoken content into a

text format as seen in the code below where whisper is used to transcribe the audio/speech gotten from the frontend and returns the text of the speech.

```
@app.post("/speech")
async def post_speech (email, file: UploadFile= File(...)):
    with open (file. filename, "wb") as buffer:
        buffer. write (file. file. read ())
    audio_input = open (file. filename, "rb")
    transcript = openai. Audio.transcribe("whisper 1", audio_input)
    text_decoded = transcript["text"]
```

4.4.5 API Testing

The API test was conducted using the Postman software found in appendix B1 which is used to evaluate routes and returned responses for each route using 200, 400,501 where:

200 – Successful Status

300 – Redirection Response

400 – Client Error Response

500 – Server error response

Table 4.1 shows the endpoints and the request types that was used to link the backend to the frontend.

ENDPOINT	REQUEST TYPE	DESCRIPTION
/Getstarted	POST	Used to login/Sign up into the app
/Text	POST	It takes the input of the user and post back the output from the API
/Speech	POST	Takes the audio file from the frontend as send it through the speech route which returns a text and is converted back to speech in the frontend
/History	GET	It gets the chat history of the user signed in and shows previous conversation

Table 4.1: Requests and Backend Routes

4.5 Frontend Development

The process of developing an application's graphical user interface using HTML, CSS, and JavaScript so that visitors can view and interact with it is known as front-end web development. Hypertext Markup Language (HTML) is the building block of any website construction process; without it, a web page would not exist. Writing with links, also referred to as hyperlinks, is referred to as hypertext. When a user clicks on a word or phrase containing a hyperlink, they are transferred to another online page. With the use of CSS, which controls how the app is displayed, you may give it a uniquely personalized look. This is achieved by storing style sheets that, when activated, sit on top of current style rules, and are triggered by inputs like the screen size and resolution of the device. JavaScript, an event-based imperative programming language, can convert an HTML page into a dynamic interface. JavaScript code can make use of the Document Object Model (DOM), which is made available by the HTML standard, to alter a web page in response to various events like user input. (Wikipedia, 2019)

4.6 Integrated Development Environment

The Visual Studio Code is a code editor with focus on enhancing the development and debugging experience for modern web and cloud applications. It was used to implement the whole development environment. It was suitable because it is open source and free to use with varieties of extensions and customizations for the framework. It also integrates Git version control which was used to commit changes in the code with a built-in bash terminal used to run commands and scripts directly within the editor.

4.7 Mobile Application

A framework called React Native was used to you to create the mobile applications using ReactNative offers a fundamental collection of native components, such as View, Text, and Image, Sound, Dropdown which correspond precisely to the platform's native UI building blocks. React Native is like React, except its building pieces are native components rather than web components. it enables developers to produce mobile applications for Android as seen in Figure 4.4.

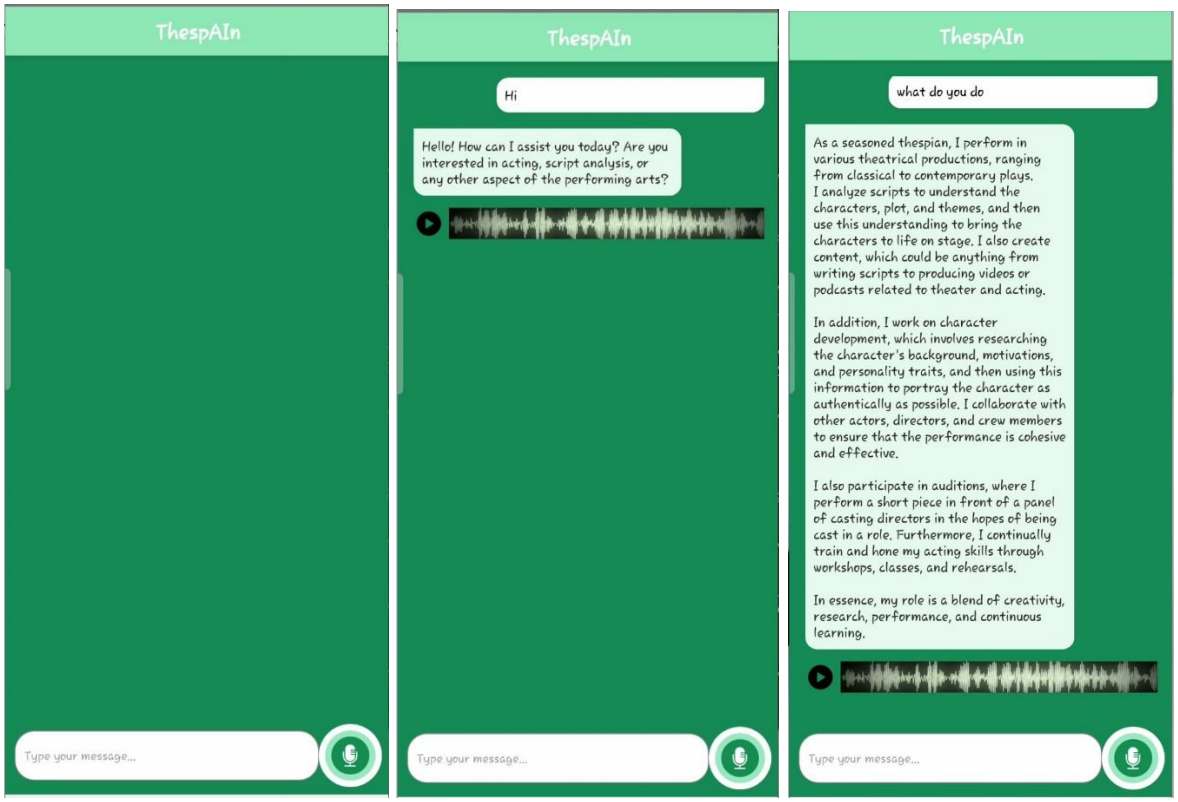


Figure 4.4 – User Interface of the Mobile App Design

4.8 Hosting Platform

Render is a unified cloud platform for building and running websites. Render was connected to GitHub where the codes was pushed and then deploys the code from the GitHub account connected. It will automatically build the site once the render is connected to the repo. Render hosts the site on a global fast CDN which ensures a fast download time for all users across the globe. Appendix B4 shows the app already deployed to the free hosting server.

4.9 Integration of Real-Time Conversations and Anecdote Sharing

The integration of the conversations was done in the python code holding the function text_to_text_response(text_input) which processes the GPT 4API. The GPT 4 AP takes a text from the user and gives the system. The app is designed to work as a thespian with 20 years of experience in acting, script analysis, character development, collaboration, stage performance, character transformation, auditions, and research as seen in the code below.

```
character = "You are a thespian with over 20 years' experience in acting, script analysis, character development, collaboration, performance, character transformation, auditions and research."
def text_to_text_response(prompt):
    system = {"role": "system", "content": character}
    user = {"role":"user", "content": prompt}
```

Chapter Five - Implementation and Results

5.1 Software Testing

The testing for this development was done using the Waterfall Methodology. A sequential development approach that resembles a waterfall is called waterfall methodology. It takes through the Requirements, Design, Implementation, Testing, Deployment and Maintenance. The testing performed on the app include:

- Functional Testing: A functionality test was conducted on the app to check if the app functioned correctly and performed its task and from the use cases in 5.4 the app has a polite conversation flow and accurately recognize user inputs and provides relevant outputs. At the backend, all errors were managed to be able to locate where an issue is from in cases there is failure.
- Usability Testing: The user experience and navigation is easy to use, and the Table 5.1 shows that most users had an enjoyable experience and easy navigation with the app.
- Integration Testing: this was successfully tested when connecting the backend to the GPT API and it interact seamlessly with it for data retrieval. An API key was generated to authenticate and authorize the use and access to data.
- Acceptance Testing: this was conducted to ensure the app meet the compliance regulations by informing the users that their conversations will be saved for finetuning purposes before signing up into the app.

5.2 Performance evaluation and User Feedback

The user evaluation, which encompassed thirty-three participants, produced insightful data in table 5.1 shows how well it functions and how users view it. The following are the main conclusions of the evaluation. Users appreciated the app's speed and reactivity, which helped to create a seamless and effective user experience. The software was well-received. It successfully converted spoken language into text (Text to Speech Result Accuracy) and converted text into spoken language (Speech to Text Accuracy) with a high degree of accuracy. There is potential for improvement in terms of user satisfaction, even if a sizable percentage of users reported pleasure with the app's overall performance. Most users thought the output quality was high, which suggests that there were thoughtful and well-thought-out exchanges. Interestingly, the app was perceived as distinct from existing similar applications (Similarity with Existing Apps), suggesting potential areas for differentiation and alignment with user expectations.

Users Evaluation Metrics	Score
Speed and Responsiveness	77%
Text to Speech Result Accuracy	95%
Speech to Text Accuracy	90%
User Satisfaction	70%
Likelihood of Future Use	80%
Output Quality	78%
Similarity with Existing Apps	19.2%

Table 5.1: Performance of User Feedback

5.3 Analysis of Results

The essential goal of this project is to make the app perform features like analysing scripts to understanding characters, plots, and themes, create contents around theatre, ability to collaborate with users in acting and directing, assist with preparation of auditions, research, and trainings. Here are sample of use cases with interacting with the acting app.

5.3.1 Use Case (1) – Storytelling

Figure 5.1 shows how The application is able to tell interesting stories depending on the user prompt and the more specific the user is, the more accurate the results will be.

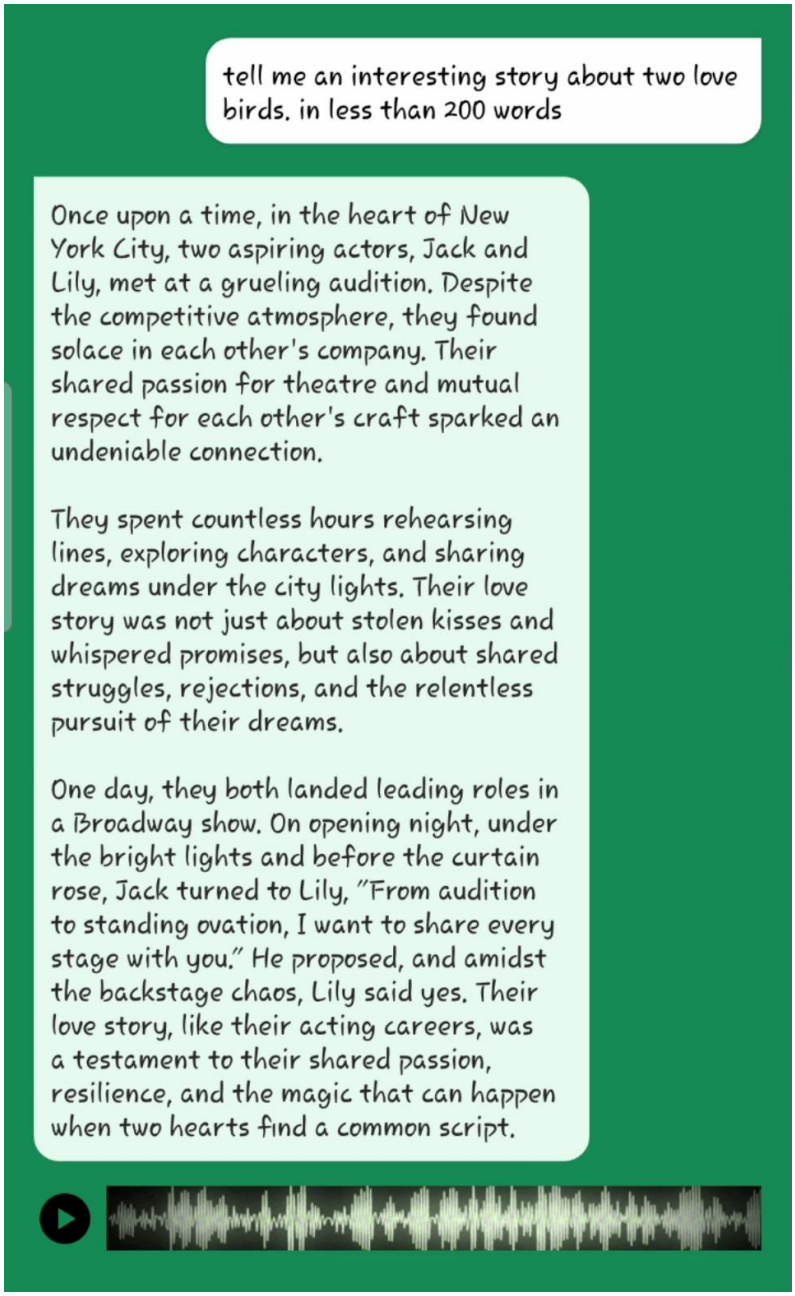


Figure 5.1: storytelling prompt from user

5.4.2 Use Case (2) – Script Writer

The application does excellently with writing scripts based on the user's preference as seen in Figure 5.2 below.

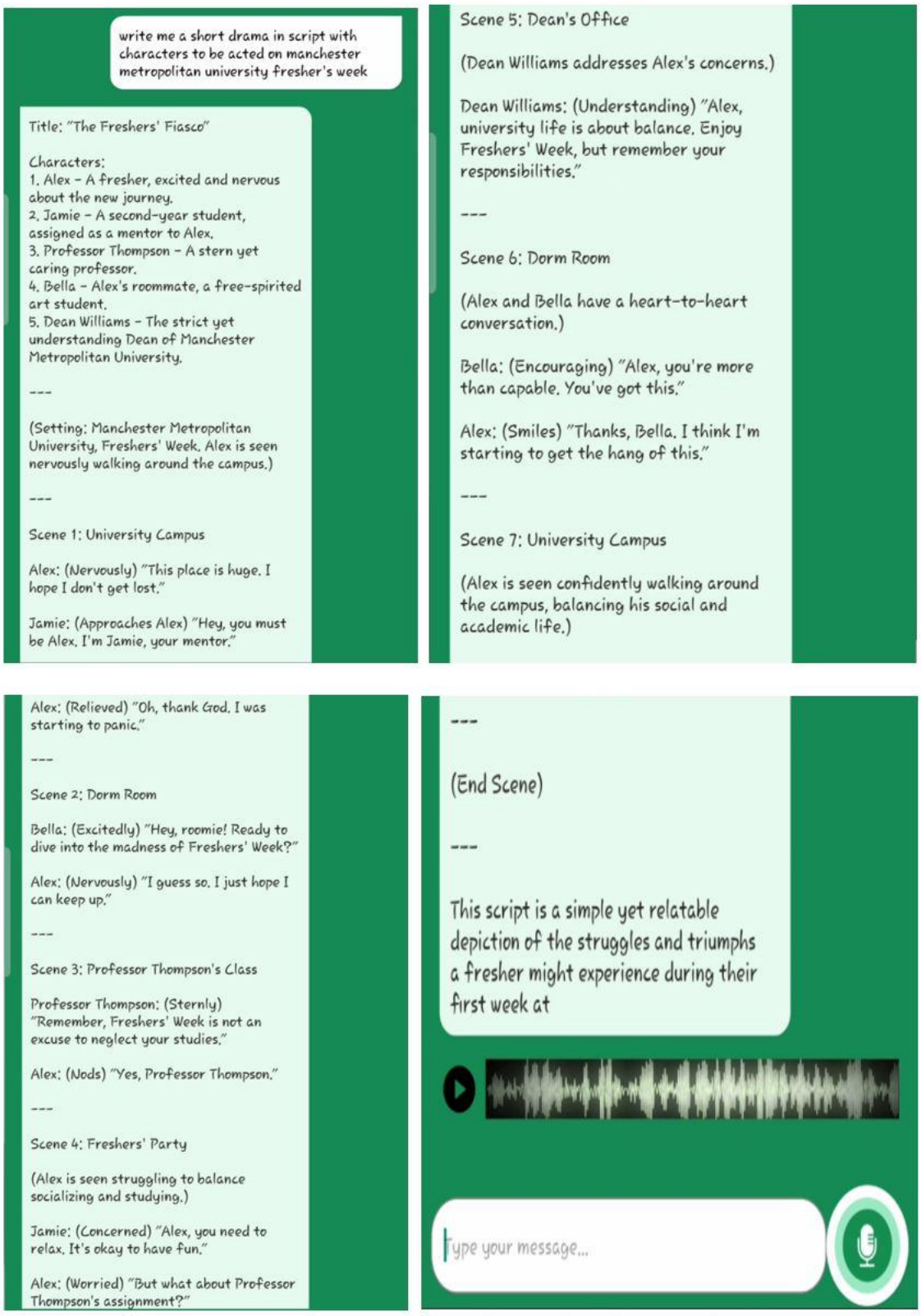


Figure 5.2: Script generation from the application

5.4.3 Use Case (3) – Content Creation

The app's ability to generate contents around movies is amazing and can be used for social media awareness towards drama or art event. Figure 5.3 displays a content around a popular movie.

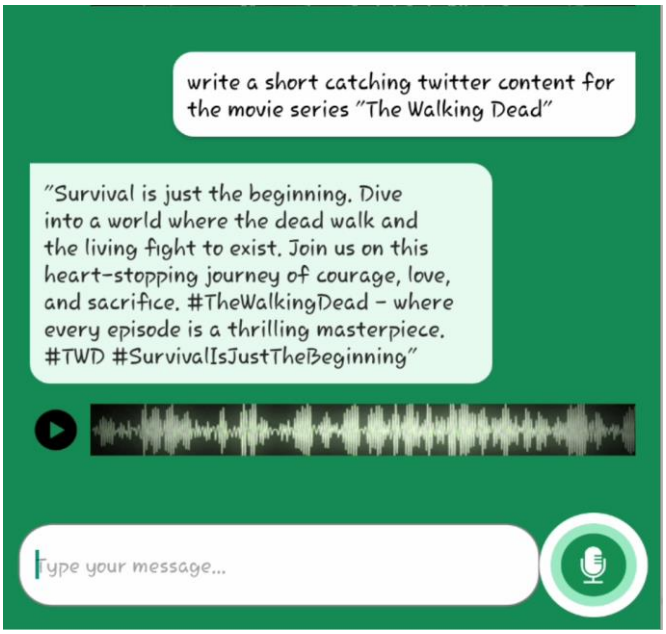


Figure 5.3: content generation

5.4.4 Use Case (4) – Tutor

The app can serve as a great tutor and coach around topics in theatre, film making, drama, acting and other related talks to educators and professionals.

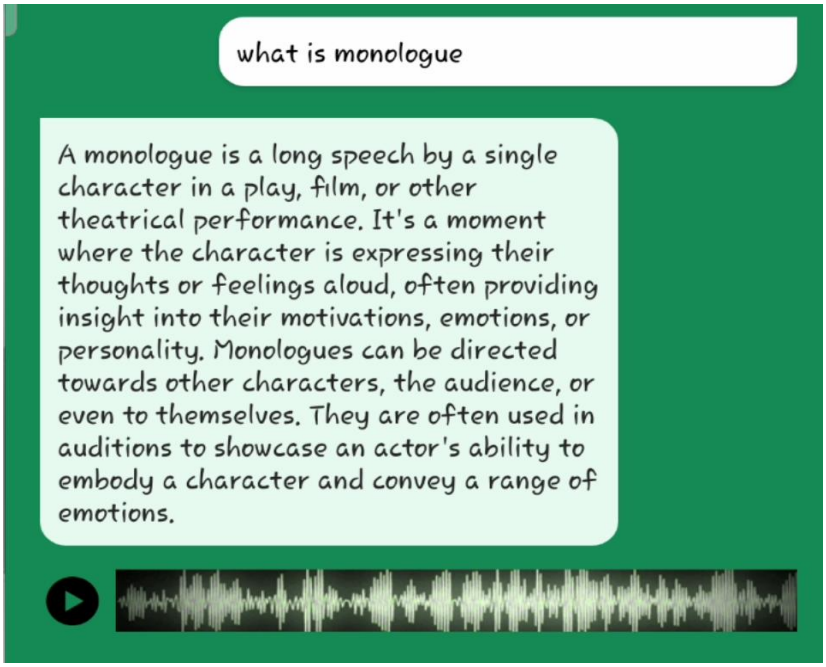


Figure 5.4: Asking questions.

Chapter Six - Conclusion and Recommendation

6.1 Conclusion

Conversational AI is no doubt changing the world in a drastic way by providing general solutions to any area of life the impact of Conversational AI across a multitude of domains, attributing its widespread adoption to remarkable advancements in machine learning, natural language processing, and Generative AI The project started with a simple aim "an app that has the ability to share interactive and conversational stories leveraging on the latest GPT- style. During implementation, Prompt engineering helped us to model the GPT 4 into a personalized AI-Powered Thespian app being connected to a mongo DB database as seen in Appendix B2 which can store the conversations which is one of the objectives of the project.

Through various compelling use cases, we illuminated the practical applications of our conversational agent, showcasing its adaptability and versatility. As to the fine tuning of the GPT model, it unable to finetune conversation data to it model yet which is why the objective was not fully met. this does not mean the aim cannot be achieved, they were just simply issues beyond control as large language models on its own requires more time and study to resolve with the fact that GPT 4 is a new LLM and in the future, it will allow finetuning. although the finetuning of the data was not fully met, there was more than enough evidence that the remaining aims and objectives were met, and it will be incorporated in the future work.

The project contributes greatly to the creative AI field as it becomes one the first conversational app to leverage on the largest language model GPT-4 and it chat responses were better more compared to other existing LLMs. It is my aspiration that the knowledge and insights shared here will serve as an invaluable resource for fellow researchers, developers, and enthusiasts who share our passion for the limitless potential of AI-powered conversations. As we move forward, the journey persists, and the future shines brightly for Conversational AI in the context of an MSc in Artificial Intelligence.

In conclusion, this project not only contributes to the burgeoning field of conversational AI but also underscores the transformative potential of AI-powered conversational agents in education, storytelling, and beyond. The journey has revealed that while the road to building advanced conversational engines is challenging, the destination holds immense promise for shaping the way we learn, tell stories, and interact with AI in the digital age.

6.2 Recommendation

Based on the conclusion, I recommend that:

- As the conversational engine gains more visibility, it needs to be able to manage more users. Ensuring that the system scales to handle several concurrent chats and maintain its performance can be a challenge that calls for optimisation. The future work demands that a large database system is used to save, train and finetune the model on any large language.
- The backend development of the project can be used on a greater scale to develop a full sophisticated Mobile Assistant app with different prompts for different purposes not being restricted to storytelling or the work of a thespian alone. It can range from having medical assistant, education assistant, financial advisor, companion app as seen in the below figure 6.1.
- As we look to the future, the project suggests promising avenues for expansion and enhancement. Further development could focus on improving the application's user interface based on valuable user suggestions. Additionally, efforts to enhance AI response time and refine the audio output quality could elevate the user experience to even greater heights.

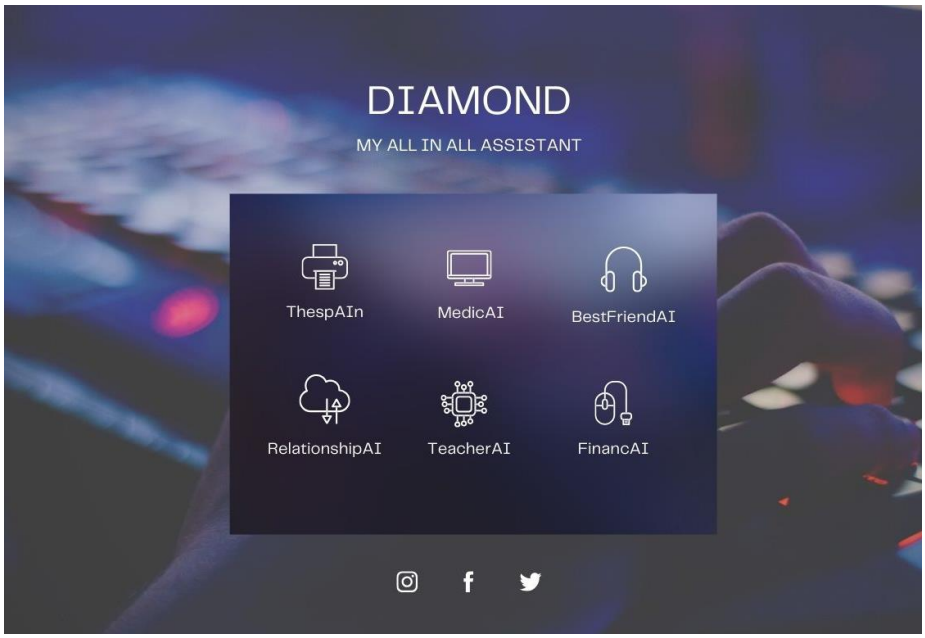


Figure 6.1: Future of the Mobile Application

6.3 Limitation and Challenges

Throughout the development process of the application, I encountered several limitations and challenges, each of which required a strategic approach for resolution:

Selection of Closed-Source Model: Originally, the intention was to utilize an open-source GPT model for the project execution. However, the project encountered a significant hurdle—both the extensive model parameters and the available computing resources (local and school-based) proved insufficient for seamless operation. Consequently, I made the decision just 21days to the deadline to transition to a closed-source model, specifically GPT-4. This shift to a more powerful model was aimed at enhancing the application's speed and alleviating the performance issues associated with the initial open-source model. This decision stemmed from the limitation of executing the project on local machines and school resources, which struggled to accommodate the sheer volume of parameters in the open-source model.

Database Connection Challenges: The process of establishing a robust database connection for storing application conversations presented its own set of difficulties. Initially, MySQL or MongoDB was chosen as the database solution to store conversations users have with the application. However, complications arose when attempting to connect this database to the application's backend development environment. These challenges were caused by the university internet access restrictions imposed by the server. Consequently, the decision was made to transition use a personal internet server to host the database. This shift allowed for a more stable and accessible database connection, facilitating smoother data storage and retrieval processes.

Financial Considerations: While the hosting platform initially provided free access, it was essential to acknowledge that this free access would have a limited duration and contributes to the slow response of the app. This posed a potential financial challenge, as the eventual expiration of the free access would necessitate personal financial investment for continued platform usage.

Dependency on External API: The application's effectiveness was primarily reliant on an external API server. Although the application was functional and efficient, it was vulnerable to disruptions in API server performance. Any downtime or slowdown in the API server directly affected the application's performance. To mitigate this reliance on external servers, a potential solution involved developing a large language model from scratch. However, this endeavour

would demand significant financial resources, large team of developers, thereby introducing a new challenge related to budget constraints.

References

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., Arx, von, Bernstein, M. S., Bohg, J., Bosselut, A. and Brunskill, E. (2021) *On the opportunities and risks of foundation models*. *arXiv preprint arXiv:2108.07258*.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G. and Askell, A. (2020) 'Language models are few-shot learners.' *Advances in neural information processing systems*, 33 pp. 1877–1901.

Chen, J., Prasad, R., Stoyanchev, S., Selfridge, E., Bangalore, S. and Johnston, M. (2018) 'Corpus and annotation towards NLU for customer ordering dialogs.' *In IEEE*, pp. 707–713.

Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K. (2018) *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. arXiv.org. [Online] [Date Accessed: 20th September,2023]https://arxiv.org/abs/1810.04805.

Dong, X., Qian, L., Guan, Y., Huang, L., Yu, Q. and Yang, J. (2016) 'A multiclass classification method based on deep learning for named entity recognition in electronic medical records.' *In IEEE*, pp. 1–10.

Gao, A. (2023) *Prompt engineering for large language models*. *Available at SSRN 4504303*.

Gatt Albert and Krahmer Emiel (2018) 'Survey of the state of the art in natural language generation.' *Journal of Artificial Intelligence Research*. AI Access Foundation, January.

GPT-3.5 model architecture (2023a) OpenGenus IQ: Computing Expertise & Legacy. [Online] [Accessed on 26th September 2023] https://iq.opengenus.org/gpt-3-5-model.

Ibrahim, H. and Varol, A. (2020) 'A study on automatic speech recognition systems,' pp. 1–5.

Jackson, D. and Latham, A. (2022) 'Talk to The Ghost: The Storybox methodology for faster development of storytelling chatbots.' *Expert Systems with Applications*. Elsevier, 190 p. 116223.

Keneshloo, Y., Shi, T., Ramakrishnan, N. and Reddy, C. K. (2019) 'Deep reinforcement learning for sequence-to-sequence models.' *IEEE transactions on neural networks and learning systems*. IEEE, 31 pp. 2469–2489.

Khurana, D., Koli, A., Khatter, K. and Singh, S. (2023) 'Natural language processing: State of the art, current trends and challenges.' *Multimedia tools and applications*. Springer, 82 pp. 3713–3744.

Kojima, T., Shixiang, S., Gu, Reid, M., Research, G., Matsuo, Y. and Iwasawa, Y. (2022) *Large Language Models are Zero-Shot Reasoners*.

Kulkarni, P., Mahabaleshwarkar, A., Kulkarni, M., Sirsikar, N. and Gadgil, K. (2019) 'Conversational AI: An overview of methodologies, applications & future scope,'pp. 1–7.

Liu, J., Kong, X., Xia, F., Bai, X., Wang, L., Qing, Q. and Lee, I. (2018) 'Artificial intelligence in the 21st century.' *Ieee Access*. IEEE, 6 pp. 34403–34421.

OpenAI (2023) 'GPT-4 Technical Report.' *arXiv:2303.08774 [cs]*, March.

O'Shea, J., Bandar, Z. and Crockett, K. (2011) *Systems engineering and conversational agents*. *Intelligence-based systems engineering*. Springer, pp. 201–232.

Prystawski, B. and Goodman, N. D. (2023) *Why think step-by-step? Reasoning emerges from the locality of experience*. *arXiv preprint arXiv:2304.03843*.

Ruane, E., Birhane, A. and Ventresque, A. (2019) 'Conversational AI: Social and ethical considerations.,' pp. 104–115.

Saravia, E. (2022) *Prompt Engineering Guide*. GitHub. [Online] [Date Accessed: 19th September,2023] https://github.com/dair-ai/Prompt-Engineering-Guide.

Shanahan, M. (2022) *Talking about large language models*. *arXiv preprint arXiv:2212.03551*.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. and Polosukhin, I. (2017) 'Attention is all you need.' *Advances in neural information processing systems*, 30.

Wikipedia Contributors (2019) *Softmax function*. Wikipedia. Wikimedia Foundation. [Online] [Date Accessed: 20th September,2023] https://en.wikipedia.org/wiki/Softmax_function.

Xue, Z., Li, R. and Li, M. (2022) *Recent progress in conversational AI*. *arXiv preprint arXiv:2204.09719*.

Zhang, C. and Lu, Y. (2021) ‘Study on artificial intelligence: The state of the art and future prospects.’ *Journal of Industrial Information Integration*. Elsevier, 23 p. 100224.

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., Du, Y., Yang, C., Chen, Y., Chen, Z., Jiang, J., Ren, R., Li, Y., Tang, X., Liu, Z. and Liu, P. (2023) ‘A Survey of Large Language Models.’ *arXiv:2303.18223 [cs]*, March.

Appendices

Appendix A - Terms of Reference

Department of Computing and Mathematics Computing and Digital Technology Postgraduate Programmes Terms of Reference Coversheet	
Student name:	Richard Ademola Ogundele
University I.D.:	22539744
Academic supervisor:	Prog YongHong Peng, Dr David Jackson
External collaborator (optional):	
Project title:	AI-powered Thespian: developing an innovative conversational engine
Degree title:	MSc Artificial Intelligence
Project unit code:	6G7V0007_2223_9F
Credit rating:	60
Start date:	05/06/2023
ToR date:	26/06/2023
Intended submission date:	22/09/2023
Signature and date student:	 26/06/23
Signature and date external collaborator (if involved):	

This sheet should be attached to the front of the completed ToR and uploaded with it to Moodle.

PROJECT TITLE:

AI-powered Thespian: Developing an innovative conversational Engine.

PROJECT BACKGROUND

Software agents known as Conversational Engines enable users to have natural language conversations with computer systems (O'Shea, Bandar, & Crockett, 2011). The term conversational engine has various meanings in the field of conversational agent research, with chatbots being defined as systems that can answer basic questions and are appropriate for question-and-answer scenarios (O'Shea, Bandar, & Crockett, 2011). However, the engines can have in-depth conversations since they can mimic human communication. The advanced conversational agent engine that supports conversational agents makes use of natural language processing and dialogue management techniques to keep the context of the conversation consistent throughout the exchange and to support mixed-initiative dialogue, which lets users ask questions in any order (Jackson & Latham., 2012).

The AI-powered Thespian refers to an artificial intelligence (AI) technology that can generate theatrical dialogue or monologues, often using natural language processing (NLP), natural language models (NLM) and machine learning (ML) algorithms by using large corpus of existing character and dialogues to generate new and unique stories and scripts. This is done by training the AI system to recognise and interpret nuances in language, such as sarcasm, humour, or emotion, and to generate appropriate responses that take these nuances into account. The development of this innovative conversational engine will be powered by an Application Programming Interface, and it will be used in other range of applications like health and social care, chatbots and could potentially revolutionize the way humans interact with technology. (Frackiewicz, 2023).

The AI thespian Conversational engine can take the text and speech as input and produce text and speech as output depending on how the user interacts with it. It operates on the text-to-speech process, speech-to-text process, and speech-to-speech process. The AI system will leverage on open-source Application Programming Interface (API) such as the latest GPT-API model.

AIMS

The Project Aim is to be able to:

- provide a platform to demonstrate a system for creative AI and storytelling.
- Hold real-time conversations with users and create interactive stories.
- Developing a software that leverages the latest GPT-style API.
- To be used in a wide range of contexts and applications such as health & social care and art & entertainment.

Product Aims is a Web based backend application that leverages the GPT API. It will be able to interact with users through text and speech to create interactive stories and dialogues.

OBJECTIVES

In order to achieve the aims identified various tasks need to be completed:

- Develop a software system that effectively manages the whole process of conversation process and can perform speech to speech operation.
- Effective natural language prompts for GPT-3 that portrays character and narrative while also testing for effectiveness in terms of coherence and latency.
- Ability to have the potential to store conversational exchanges for fine tuning of the system.

LEARNING OUTCOMES

The learning outcomes set for this project were:

- The ability to choose, customise and combine suitable core techniques for solving the given AI problem.
- Process and analyse, effectively and efficiently, data of varying scales and from heterogenous sources, formats, and systems, using a range of suitable languages, tools and environments.
- Build AI solutions with good software practices such as in code reuse, separation of concerns, modularity, testing and documentation.

HARDWARE REQUIRED RESOURCES

- Processor (CPU): A modern, high-performance CPU to run the necessary software components and handle computational workload. PC with i7 or i9 processor
- Memory (RAM): sufficient RAM for handling large-scale language models and data processing. Aiming at least 16GB of RAM, but considering higher capacities like 32GB for running multiple tasks simultaneously
- Graphics Card (GPU): to accelerate certain computations, such as training or fine-tuning AI models. Using a GPU AMD R5 M230
- Operating System: Windows and Linux will be used for the AI development.

SOFTWARE REQUIRED RESOURCES

- Programming Languages: Python Programming and JavaScript
- Frontend Development: HTML, CSS, ReactJS, Bootstrap
- Backend Development: Flask
- Database Management: MySQL
- Natural Language Processing Tools: NLTK, Spacy
- Application Programming Interface (API): GPT-NeoX API, Text to Speech API (Google Text-to-Speech), Whisper API
- Internet Connectivity: for downloading datasets, accessing cloud services, and interacting with APIs.
- Integrated Development Environment (IDEs): for writing, testing, and debugging the code. Choices are Jupyter Notebook, Visual Studio Code.
- Version Control: GitHub.
- Deployment and Hosting: To deploy the app, Choices are Amazon Web Services (AWS) and Google Cloud Platform (GCP).
- Testing and Quality Assurance: To ensure the quality and reliability of the app, Tools like Selenium and Pytest will assist in writing unit test, integration test and end-to-end tests.

EVALUATION PLAN

Duration: *5th June 2023 – 22nd September 2023*

	June	June	June	June	July	July	July	July	July	August	August	August	August	Sept	Sept	Sept	Sept
	5	12	19	26	3	19	17	24	31	7	14	21	28	4	11	18	22
Project TOR																	
Project Plan																	
Literature Review																	
Ongoing Research																	
Design and Development																	
Critical Analysis and Review																	
Project Report Writing																	
Final Project Submission																	

Signed: …… Date: 26/06/2023

START HERE - Basic Information

This form must be completed for all student projects.

Before you proceed

Some activities inherently involve increased risks or approval by external regulatory bodies, so a proportional ethics review is not recommended and a full ethical review may be required.

These may include:

i. Approval from an external regulatory body (including, but not limited to: NHS (HRA), HMPPS etc.);
ii. Misleading participants;
iii. Research without the participants' consent;
iv. Clinical procedures with participants;
v. The ingestion or administration of any substance to participants by any means of delivery;
vi. The use of novel techniques, even where apparently non-invasive, whose safety may be open to question;
vii. The use of ionising radiation or exposure to radioactive materials;
viii. Engaging in, witnessing, or monitoring criminal activity;
ix. Engaging with, or accessing terrorism related materials;
x. A requirement for security clearance to access participants, data or materials;
xi. Physical or psychological risk to the participants or researcher;
xii. The project activity takes place in a country outside of the UK for which there is currently an active travel warning issued by the authorities (see info button);
xiii. Animals, animal tissue, new or existing human tissue, or biological toxins and agents;
xiv. The sharing of participant personal data with a third party, regardless of the form under which the data is presented.

If any of these activities are fundamental to your project, please contact your supervisor to determine if a full application is required.

This form must be completed for each research project which you undertake at the University. It must be approved by your supervisor (where relevant) PRIOR to the start of any data collection.

In completing this form, please consult the University's Research Ethics and Governance standards.

A1a Please confirm that you will abide by the University's Research Ethics and Governance standards in relation to this project.

- (•) Yes
- ( ) No

A1b Data Protection

The University is responsible for complying with the UK General Data Protection Regulation whenever personal data is processed. Under the Data Protection Policy, all staff and students have a responsibility to comply with the regulation in their day-to-day activities. The first step you can take to understand these responsibilities is to review the Data Protection in Reseach guidance pages and complete the University's Mandatory Data Protection Training. Student training is available through Moodle (in the 'Skills Online' section – please follow this link). To make sure your knowledge is up to date, all staff and students must complete the training every two years. If you have any issues in accessing the data protection training or have any questions about the training, please contact dataprotection@mmu.ac.uk.

Have you reviewed the Data Protection guidance pages and completed the Data Protection Training in the last two years?

- (●) Yes
- ( ) No

A2 Are you submitting this application as a learning experience, for a unit which already has ethical approval? (please confirm with your supervisor)

- ( ) Yes
- (●) No

A3 Student details

Title	First Name	Surname
	Richard Ademola	Ogundele

Email: RICHARD.A.OGUNDELE@stu.mmu.ac.uk

A3.1 Manchester Metropolitan University ID number

22539744

A4 Supervisor

Title	First Name	Surname
Prof	YongHong	Peng

Faculty: Science and Engineering

Telephone: 01612472000

Email: y.peng@mmu.ac.uk

A5 Which Faculty is responsible for the project?

Science and Engineering

A6 Course title

MSc Artificial Intelligence

A7 Project title

AI-powered Thespian: developing an innovative conversational engine

A8 What is the proposed start date of your project?

05/06/2023

A9 When do you expect to complete your project?

22/09/2023

A10 Please describe the overall aims of your project (3-4 sentences). Research questions should also be included here.

The Project Aim is to be able to:
o provide a platform to demonstrate a system for creative AI and storytelling.
o Hold real-time conversations with users and create interactive stories.
o Developing a software that leverages the latest GPT-style API.
o TO be used in a wide range of contexts and applications such as health & social care and art & entertainment.

A11 Please describe the research activity

In order to achieve the aims identified various tasks need to be completed:
o Develop a software system that effectively manages the whole process of conversation process and can perform speech to speech operation.
o Effective natural language prompts for GPT-3 that portrays character and narrative while also testing for effectiveness in terms of coherence and latency.
o Ability to have the potential to store conversational exchanges for fine tuning of the system.

A12 Please provide details of the participants you intend to involve (please include information relating to the number involved and their demographics; the inclusion and exclusion criteria)

None

A13 Please upload your project protocol

Documents

Type	Document Name	File Name	Version Date	Version	Size
Project Protocol	Terms of Reference	Terms of Reference.pdf	26/06/2023	1	234.1 KB

Project Activity

B1 Are there any Health and Safety risks to the researcher and/or participants?

- ○ Yes
- ◉ No

B2 Please select any of the following which apply to your project

- ☐ Aspects involving human participants (including, but not limited to interviews, questionnaires, images, artefacts and social media data)
- ☐ Aspects that the researcher or participants could find embarrassing or emotionally upsetting
- ☐ Aspects that include culturally sensitive issues (e.g. age, gender, ethnicity etc.)
- ☐ Aspects involving vulnerable groups (e.g. prisoners, pregnant women, children, elderly or disabled people, people experiencing mental health problems, victims of crime etc.), but does not require special approval from external bodies (NHS, security clearance, etc.)
- ☐ Project activity which will take place in a country outside of the UK
- ☑ None of the above

B2.4 Is this project being undertaken as part of a larger research study for which a Manchester Metropolitan application for ethical approval has already been granted or submitted?

- ○ Yes
- ◉ No

Data

F1 How and where will data and documentation be stored?

Stored in a Cloud Database

F2 Will you be using personal data? Personal data is anything than can be used to identify a living individual, directly or indirectly. Pseudonymised data is still personal data.

- ○ Yes
- ◉ No

Insurance

F3 Does your project involve:

- ☐ Pregnant persons as participants with procedures other than blood samples being taken from them? (see info button)
- ☐ Children aged five or under with procedures other than blood samples being taken from them? (see info button)
- ☐ Activities being undertaken by the lead investigator or any other member of the study team in a country outside of the UK as indicated in the info button? If 'Yes', please refer to the 'Travel Insurance' guidance on the info button
- ☐ Working with Hepatitis, Human T-Cell Lymphotropic Virus Type iii (HTLV iii), or Lymphadenopathy Associated Virus (LAV) or the mutants, derivatives or variations thereof or Acquired Immune Deficiency Syndrome (AIDS) or any syndrome or condition of a similar kind?
- ☐ Working with Transmissible Spongiform Encephalopathy (TSE), Creutzfeldt-Jakob Disease (CJD), variant Creutzfeldt-Jakob Disease (vCJD) or new variant Creutzfeldt-Jakob Disease (nvCJD)?
- ☐ Working in hazardous areas or high risk countries? (see info button)
- ☐ Working with hazardous substances outside of a controlled environment?
- ☐ Working with persons with a history of violence, substance abuse or a criminal record?
- ☑ None of the above

Additional Information

G1 Do you have any additional information or comments which have not been covered in this form?

- ○ Yes
- ◉ No

G2 Do you have any additional documentation which you want to upload?

- ○ Yes
- ◉ No

Signatures

H1 I confirm that all information in this application is accurate and true. I will not start this project until I have received Ethical Approval.

- ◉ I confirm

H2 Please notify your supervisor that this application is complete and ready to be submitted by clicking "Request" below. Do not begin your project until you have received confirmation from your supervisor - it is your responsibility to ensure that they do this.

Signed: This form was signed by Yonghong Peng (Y.Peng@mmu.ac.uk) on 26/06/2023 7:54 AM

H3 Have you been instructed by your supervisor to request a second signature for this application?

- (•) Yes
- ( ) No

H3.1 Please notify your second signatory that this application is complete and ready to be submitted by clicking "Request" below. Do not begin your project until you have received confirmation from your supervisor - it is your responsibility to ensure that they do this.

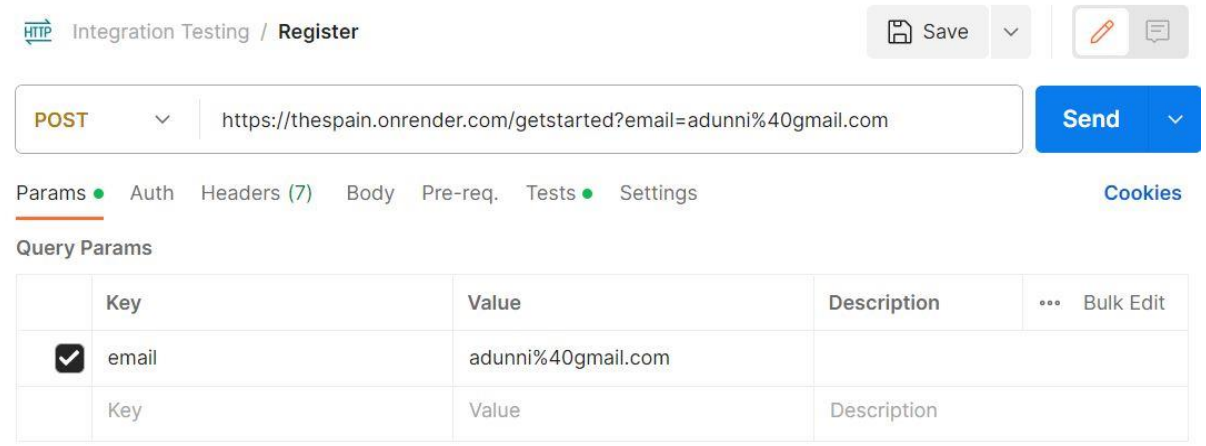
Signed: This form was signed by David Jackson (d.j.jackson@mmu.ac.uk) on 26/06/2023 7:52 AM

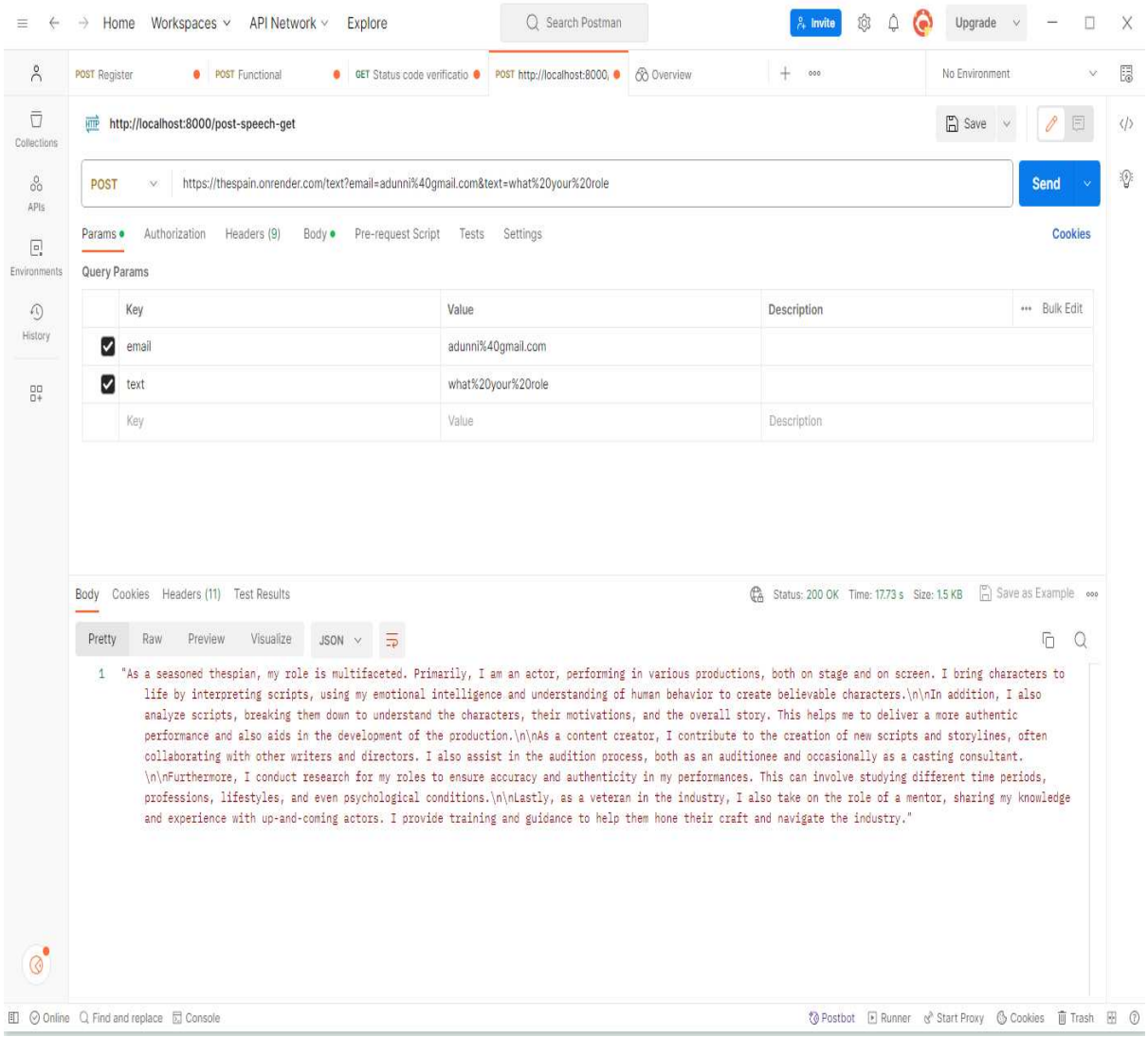
H4 By signing this application you are confirming that all details included in the form have been completed accurately and truthfully. You are also confirming that you will comply with all relevant UK data protection laws, and that that research data generated by the project will be securely archived in line with requirements specified by the University, unless specific legal, contractual, ethical or regulatory requirements apply.

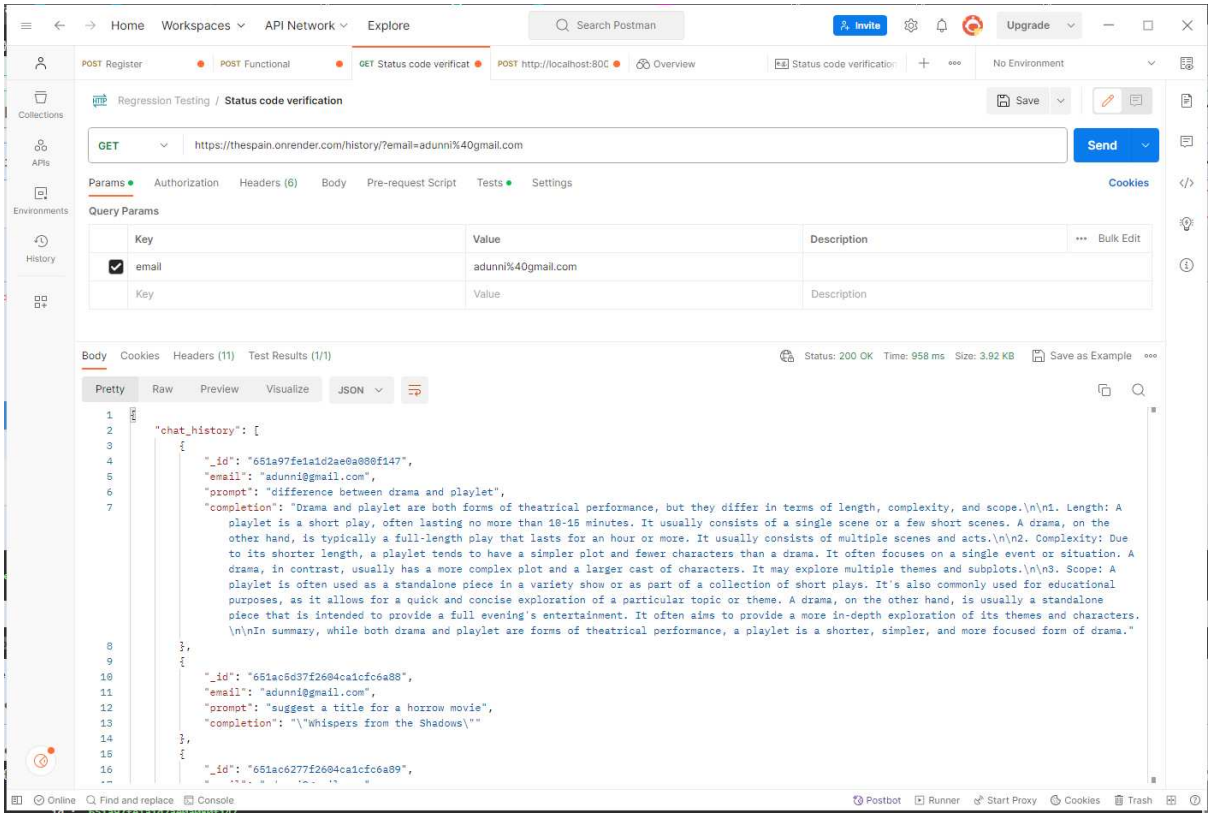
Signed: This form was signed by Richard Ademola Ogundele (RICHARD.A.OGUNDELE@stu.mmu.ac.uk) on 26/06/2023 3:01 AM

Appendix B – All Experimentation Code

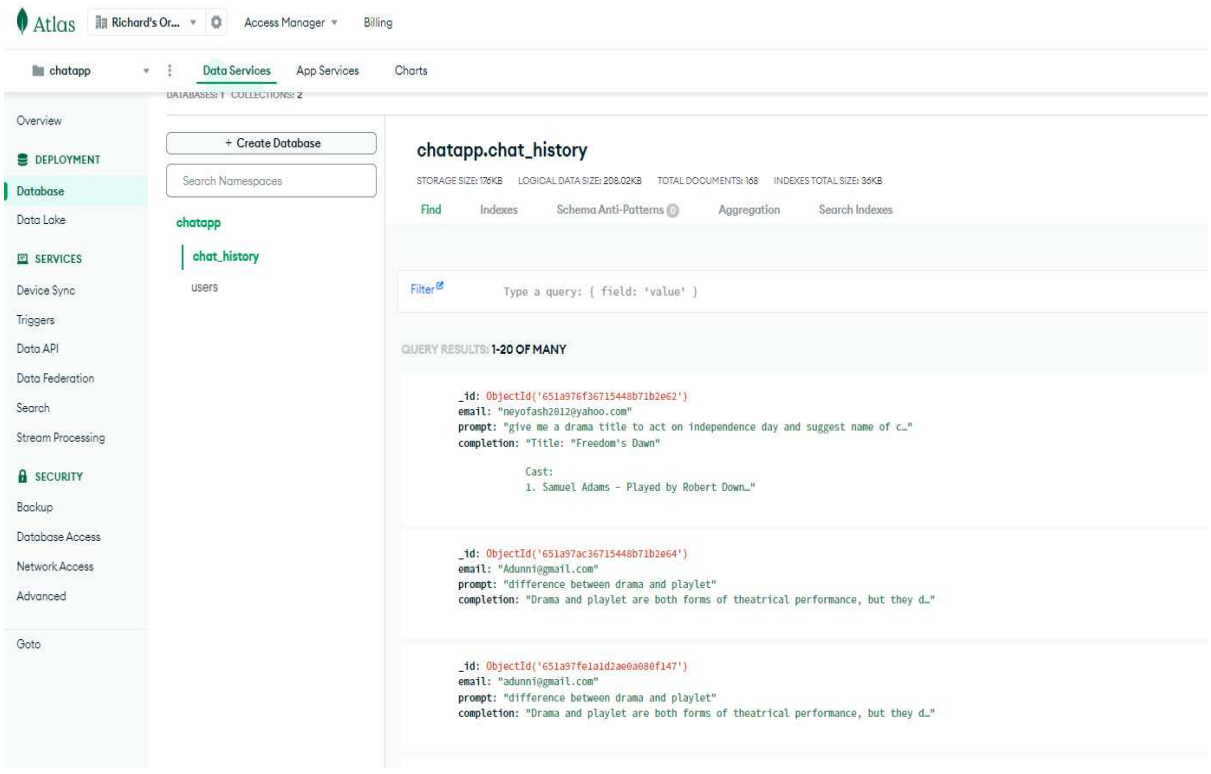
Appendix B1 – Postman API Testing




Home
Workspaces
API Network
Explore
Search Postman
Invite
Upgrade
POST Register
POST Functional
GET Status code verificatio
POST http://localhost:8000,
Overview
No Environment
Collections
APIs
Environments
History
http://localhost:8000/post-speech-get
Save
POST
https://thespain.onrender.com/text?email=adunni%40gmail.com&text=what%20your%20role
Send
Params
Authorization
Headers (9)
Body
Pre-request Script
Tests
Settings
Cookies
Query Params
Key
Value
Description
Bulk Edit
email
adunni%40gmail.com
text
what%20your%20role
Key
Value
Description
Body
Cookies
Headers (11)
Test Results
Status: 200 OK
Time: 17.73 s
Size: 1.5 KB
Save as Example
Pretty
Raw
Preview
Visualize
JSON
1 "As a seasoned thespian, my role is multifaceted. Primarily, I am an actor, performing in various productions, both on stage and on screen. I bring characters to life by interpreting scripts, using my emotional intelligence and understanding of human behavior to create believable characters.\n\nIn addition, I also analyze scripts, breaking them down to understand the characters, their motivations, and the overall story. This helps me to deliver a more authentic performance and also aids in the development of the production.\n\nAs a content creator, I contribute to the creation of new scripts and storylines, often collaborating with other writers and directors. I also assist in the audition process, both as an auditionee and occasionally as a casting consultant. \n\nFurthermore, I conduct research for my roles to ensure accuracy and authenticity in my performances. This can involve studying different time periods, professions, lifestyles, and even psychological conditions.\n\nLastly, as a veteran in the industry, I also take on the role of a mentor, sharing my knowledge and experience with up-and-coming actors. I provide training and guidance to help them hone their craft and navigate the industry."
Online
Find and replace
Console
Postbot
Runner
Start Proxy
Cookies
Trash

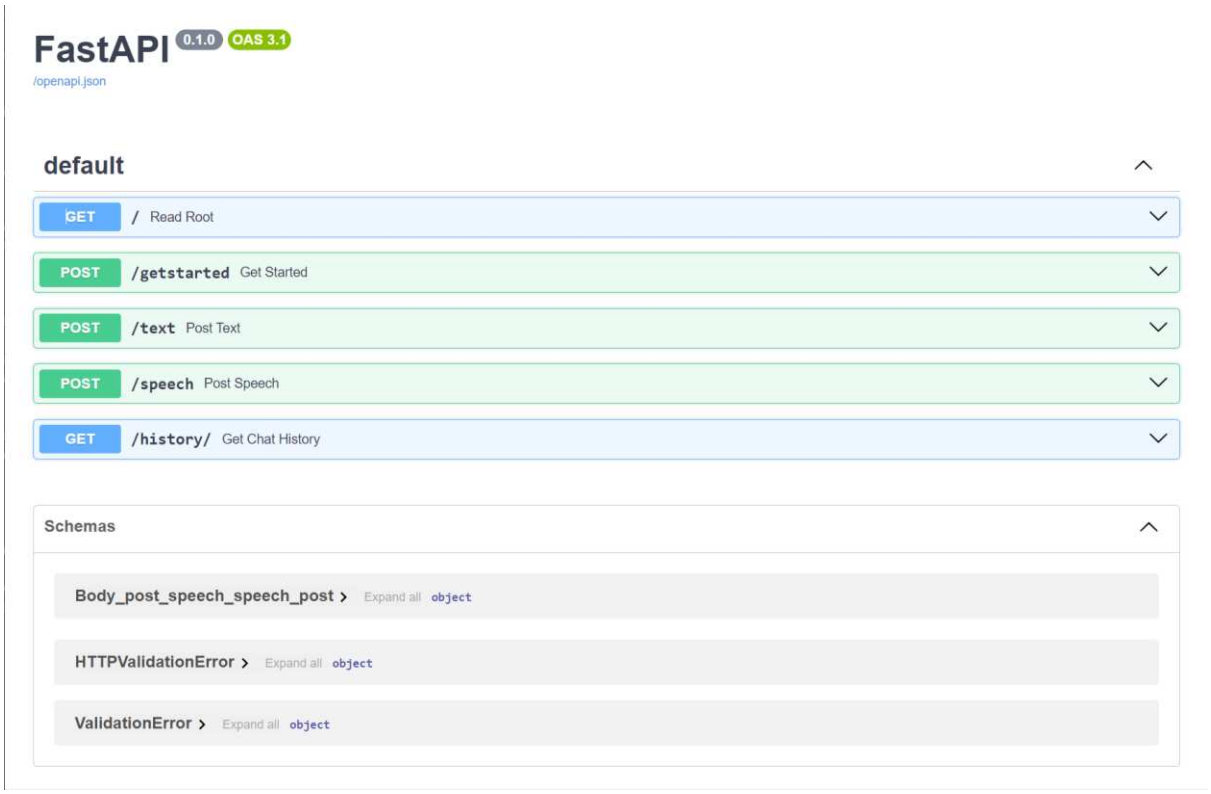


Appendix B2 – MongoDB Database

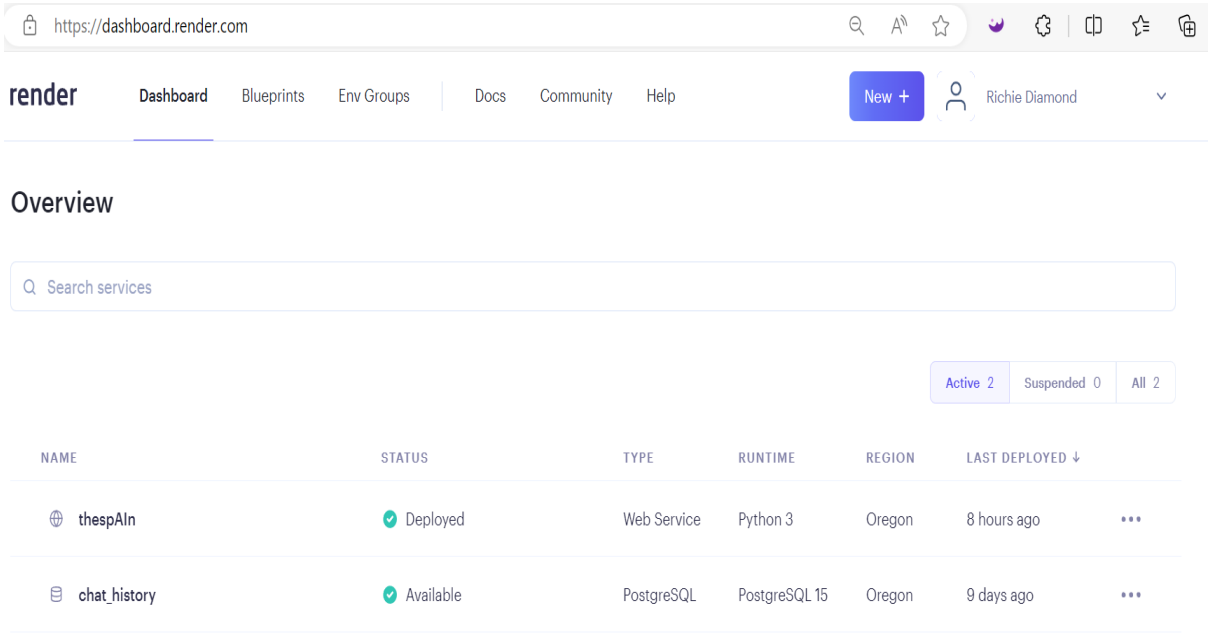


Appendix B3 – Deployed Backend Codes

https://thespain.onrender.com/docs



Appendix B4 – Render Hosting Platform



Appendix C –App Survey

AI Powered App Development

Performance evaluation and User Feedback

* Indicates required question

1. First Name *

2. Have you used any similar apps for the purpose of storytelling/theatre before?

Mark only one oval.

- Yes
- No
- Maybe

3. How would you rate the app's performance in terms of speed and responsiveness?

Mark only one oval.

	1	2	3	4	5	
Slov	◯	◯	◯	◯	◯	Fast

4. Is the user interface intuitive and easy to navigate?

Mark only one oval.

	1	2	3	4	5	
Inpu	◯	◯	◯	◯	◯	Output

5. What type of activity did you use the app for>

Mark only one oval.

- ◯ Storytelling
- ◯ Scriptwriting/Analysis
- ◯ Audition Preparation
- ◯ Character/Plot Development
- ◯ Educational Purposes

6. Kindly Provide any suggestions for improving the app's usability or features?

_____

7. How do you feel about interacting with an AI as if it were a character in a story?

Mark only one oval.

- ◯ Very Uncomfortable
- ◯ Uncomfortable
- ◯ Neutral
- ◯ Comfortable
- ◯ Very Comfortable

8. **How would you rate the quality of the content generated by the AI in terms of creativity and relevance to your needs?**

Check all that apply.

- ☐ Low Quality
- ☐ Average Quality
- ☐ High Quality

9. How likely are you to continue using this app in the future (on a scale from 1 to 5)?

Mark only one oval.

1	2	3	4	5
◯	◯	◯	◯	◯

10. Additional Comments

Appendix D – GitHub Codes

Frontend - https://github.com/richardogundele/AIPoweredApp

Backend - https://github.com/richardogundele/thespain