****

Generative AI Voice-Enabled Chatbot for the University of Dundee Website

A dissertation submitted for the Degree of Master of Science

MSc in Data Science and Engineering

School of Science and Engineering

University of Dundee

5th April 2024

**Table of Contents**

[**1.0 Introduction** 5](#_Toc163204667)

[**1.1 Problem Statement** 6](#_Toc163204668)

[**1.2 Aim and Objectives** 6](#_Toc163204669)

[**1.3 Project Outline** 6](#_Toc163204670)

[**2.0 Literature Review** 7](#_Toc163204671)

[**2.1 Introduction to Generative AI in Educational Settings** 7](#_Toc163204672)

[**2.2 AI-Driven Chatbots in University Websites** 7](#_Toc163204673)

[**2.3 Voice-Enabled Chatbot Interface** 8](#_Toc163204674)

[**2.4 Natural Language Processing** 8](#_Toc163204675)

[**2.5 Voice-Enabled Chatbots** 8](#_Toc163204676)

[2.5.1 Speech-to-text 9](#_Toc163204677)

[2.5.2 Text-to-Speech 9](#_Toc163204678)

[**2.6 User Experience and Accessibility** 10](#_Toc163204679)

[**2.7 Literature Gap** 10](#_Toc163204680)

[**2.8 Summary** 11](#_Toc163204681)

[**3.0 Specification** 12](#_Toc163204682)

[**3.1 Requirements** 12](#_Toc163204683)

[3.1.1 Functional Requirement 12](#_Toc163204684)

[3.1.2 Non-Functional Requirements 12](#_Toc163204685)

[**3.2 Project Plan** 13](#_Toc163204686)

[3.2.1 Project Time Line 13](#_Toc163204687)

[3.2.2 Justification of Methodology 13](#_Toc163204688)

[**4.0 Design** 14](#_Toc163204689)

[**4.1 Main Design Choices** 14](#_Toc163204690)

[4.1.1 Programming Languages 14](#_Toc163204691)

[4.1.2 Natural Language Processing with Open AI 14](#_Toc163204692)

[4.1.3 Speech-to-Text with Assembly AI 14](#_Toc163204693)

[4.1.4 Text-to-Speech with Google Cloud 14](#_Toc163204694)

[4.1.5 Backend Web Framework FastAPI 15](#_Toc163204695)

[**4.2 Unified Modeling Language (UML)** 15](#_Toc163204696)

[4.2.1 Use Case Diagram 15](#_Toc163204697)

[4.2.2 Sequence Diagram 16](#_Toc163204698)

[4.2.3 component Diagram 17](#_Toc163204699)

[**5.0 Implementation** 18](#_Toc163204700)

[**5.1 System Architecture** 18](#_Toc163204701)

[**5.3 Phase-Based Development** 18](#_Toc163204702)

[5.3.1 Phase 1: Open AI Integration 18](#_Toc163204703)

[5.3.3 Phase 3: Backend Development 19](#_Toc163204704)

[5.3.4 Phase 4: Frontend Development 21](#_Toc163204705)

[5.3.5 Phase 4: Finetuning Job 21](#_Toc163204706)

[**6.0 Evaluation and Testing** 22](#_Toc163204707)

[**6.1 Results** 22](#_Toc163204708)

[**7.0 Description of the Final Product** 23](#_Toc163204709)

[**7.1 Interactive User Interface** 23](#_Toc163204710)

[**7.2 Advanced Core Functionalities** 23](#_Toc163204711)

[**7.3 Accessibility and Performance** 23](#_Toc163204712)

[**8.0 Appraisal** 25](#_Toc163204713)

[**8.1 Legal, Social, Ethical and Professional Issues** 25](#_Toc163204714)

[8.1.1 Legal Issues 25](#_Toc163204715)

[8.1.2 Social Issues 25](#_Toc163204716)

[8.1.3 Ethical Issues 26](#_Toc163204717)

[8.1.4 Professional Issues 26](#_Toc163204718)

[**9.0 Summary and Conclusion** 27](#_Toc163204719)

[**10 Future Work** 28](#_Toc163204720)

[**11 References** 29](#_Toc163204721)

[**12 Appendices** 34](#_Toc163204722)

**List of Figures**

Figure 4.1: Use Case Diagram……………………………….…………………………………………………….…15

Figure 4.2: Sequence Diagram…………………………………..……………………………………………………16

Figure 4.3: Component Diagram……………………………………….…………………………………………….17

Figure 5.1: System Architecture………………………….…………………………………………………………..18

Figure 5.2: Code Snippet Open AI Integration……………………………………………………………………….19

Figure 5.3: Code Snippet Chatbot Response Post Request………………………………………………….………..20

Figure 5.4: Code Snippet Assembly AI Transcription Post Request……………………….…………………………20

Figure 5.5: Fine-Tune Job…………………………………………………………………………………………….21

Figure 7.1: Final Product……………………………………………………………………………………………..24

**Executive Summary**

In the evolving society, where individuals are becoming more dependent on technology, this dependency leads to a demand for digital resources to make life easier. The Generative AI Voice Enabled Chatbot is designed to transform the way individuals interact with digital platforms. This project was initiated to tackle a few challenges with information accessibility and digital navigation, by implementing an intuitive, interactive platform on the University of Dundee website. This project aims to enhance online user experience, by providing a platform for interaction and information retrieval. By integrating advanced AI technologies such as Open AI GPT 3.5 for natural language processing, Assembly language for real-time transcription, speech-to-text conversion, and voice recognition, and Google Cloud text-to-speech API for voice responses, the chatbot offers a dynamic and responsive platform for users to easily retrieve information, navigate services and engage with the university’s community and resources, through text and voice input. The project was guided by core objectives to enhance digital user experience, all while adhering to set standards and requirements. The development process for this project adopted Agile methodology, leveraging Agile’s quality focus and iterative development to design and build the chatbot functionalities. This project also outlines in detail the development journey, key challenges faced during the implementation process, final product, different approaches, and recommendations for future studies. The successful outcome of this project indicates not only enhancing the usability and accessibility of the University of Dundee website but also setting a standard for problem-solving with AI in the educational sector.

# **1.0 Introduction**

In the dynamic domain of educational technology. It is only ideal for any educational institution to pursue innovative and promising technologies, that can redefine the communication paradigms in our community while also catering to individuals with diverse needs. Traditional communication methods for interaction within universities, often struggle to meet the expectations of digitally savvy students and staff. This can lead to potential information gaps and a lack of engagement. Consequently, there is a pressing need for a solution that transcends conventional communication barriers, offering an intuitive, and inclusive platform for engagement [72]. The Generative AI Voice-Enabled Chatbot aims to address these issues by providing a versatile and responsive communication tool. However, the endeavors to bring this solution to life, are not without obstacles, including the provision of a seamless, intuitive user experience ensuring accuracy and maintaining a high level of user engagement [75]. This project seeks to overcome these hurdles, setting a new standard for digital communication within the educational community. This report delves into the creation of a Generative AI Voice-Enabled Chatbot, designed to promote an interactive and accessible environment for both students and staff, by leveraging the powerful capabilities of various sophisticated technologies like Python, Fast API for backend development, OpenAI's GPT 3.5 for natural language processing, Assembly AI for accurate speech-to-text conversion, and Google Cloud Text-to-Speech API for generating natural-sounding voice responses. The initiative of this project is driven by the goal of making information and services more accessible and interactive, catering to a diverse audience. The chatbot is designed to be contextually relevant, user-friendly, and engaging.

## **1.1 Problem Statement**

Traditional information retrieval and navigation methods on the University of Dundee website are not meeting the evolving expectations of digitally savvy users, leading to potential information gaps and challenges with accessing information in real time. This situation facilitates the development of an innovative solution, that transcends conventional communication barriers, offering an intuitive platform for interaction with the power of Generative AI.

## **1.2 Aim and Objectives**

The aim of implementing this chatbot is to enhance user experience, accessibility, and engagement by providing a seamless platform for interaction and information retrieval.

**Objectives**

* Improve user experience, and enhance ease of navigation and accessibility by implementing a user-friendly chatbot interface.
* Facilitate information retrieval by providing quick and accurate responses to user inquiries, allowing students, staff, and visitors to easily access relevant information about the university.
* Promote user engagement with the university website by providing interactive features on the chatbot, to encourage active engagement from users, ensuring the chatbot serves as a dynamic tool for exploring the services and resources the university has to offer.

## **1.3 Project Outline**

* Section 1: includes, introduction, problem statement, aims, and objectives.
* Section 2: discuss in detail, a background study on technologies used, a review of related work, and, literature gap
* Section 3: outlines the requirements functional and non-functional, project plan, timeline, and methodology.
* Section 4: discusses the overall design, design choices and tradeoffs, and UML diagrams
* Section 5: gives a detailed explanation of the implementation process, system architecture, and phase-based development
* Section 6: includes evaluation and testing, interview analysis, results and user feedback
* Section 7: describes the final product, its design, key functionalities, and characteristics
* Section 8 includes appraisal, which elaborates on different approaches for future studies
* Section 9: overall summary and conclusion
* Section 10: recommendations for future works.

# **2.0 Literature Review**

## **2.1 Introduction to Generative AI in Educational Settings**

Characterized by the rapid emergence of diverse innovations and technology, the contemporary world has seen their use across several sectors such as medicine, finance, and even education [1]. According to Haleem et al. [2], education, in particular, has experienced the rapid influx and integration of several learning technologies such as teaching technologies, learning technologies, and interactive technologies. However, the increased need for students to utilize and tap into enhanced communication techniques and approaches has stirred up the need to incorporate useful technologies that bridge such gaps, as traditional communication methods such as FAQ pages, emails, etc. appear to be no longer effective and foster engagement among students, teachers, and the entire university community. Thus, the need for modern technologies which offer students a more innovative and novel means of communication, appealing to their tech savviness and addressing unique needs of special individuals such as those with impairments and intellectual disabilities are quite apparent [3], [4]. Amidst a variety of potentially viable options, generative AI, AI-driven chatbots and voice-enabled chatbot interfaces stand out as befitting solutions, offering several useful features and appealing to different categories of learners [5], [6], [7]. The review is structured into the following sections:

## **2.2 AI-Driven Chatbots in University Websites**

Integrating AI-driven chatbots into university websites signifies a huge leap in digital communication processes in educational communities. The chatbots exploit artificial intelligence (AI), notably, Natural Language Processing, and employ the use of machine learning to provide assistive functions such as providing answers to questions, and optimizing administrative functions [8], [9]. This may include record management, curriculum development, etc. [10]. The advantages have been highlighted by various literature. For instance, Haugeland et al. [11] noted that AI-driven chatbots improve user experiences by accurately providing a variety of responses to prompts and queries, thereby significantly reducing the workload burden on administrative staff and ultimately boosting their efficiency [12]. Additionally, Panda and Chakravarty [13] opined that chatbots considerably facilitate user accessibility, giving room for a wide range of unique needs and preferences which otherwise might not be catered for by more crude methods.

However, despite the numerous advantages offered by incorporating AI-driven chatbots into university websites, certain studies revealed that a variety of limitations might be encountered. For instance, Mah et al. [14] explained that one of the most pressing issues is the need for a sophisticated Natural Language Processing system algorithm capable of a high level of comprehension of a wide variety of prompts and queries which users may provide. Caldarini et al. [15] noted that such sophisticated systems are tedious to use and consume an ample amount of time. Furthermore, the overall effectiveness of chatbots requires progressive machine learning and constant software updates to address the ever-evolving nature of human needs and preferences [16]. Thus, continuous adaptation is a necessity for AI-driven chatbots to be effective and fulfilling to a high level.

Thus, while AI-driven chatbots may offer a wide range of applications in educational settings, particularly university websites, the literature reviewed also underscore crucial limitations which must be considered. Although certain positives such as improved user satisfaction and a boost in efficiency of operations, certain obstacles exist such as the gap in long-term functionality, notably in the paradigm of the constant evolution of the nature of human queries and needs [11], [16], ultimately impacting on the ability of the AI systems to effectively address and provide more coherent and usable responses. In addition, the extent to which the impacts of AI-driven chatbots on user engagement is observable appears to be dependent on several factors such as the unique chatbot design, the user interface, and organizational context [17]. For instance, Natalia et al. [18] noted that chatbots that utilize sophisticated intuitive interfaces and strategic engagement tools have a higher capacity to foster better user experience as opposed to chatbots that employ more stringent operational models which result in more limited functionality.

Thus, AI-driven chatbots possess tremendous capabilities for improving communication on university websites. However, a successful systems integration is contingent on thorough assessment and consideration of the technicalities involved, the design and organizational factors and wider context. Addressing these factors potentially leads to a successful incorporation.

## **2.3 Voice-Enabled Chatbot Interface**

Incorporating voice-enabled chatbots into the university website reflects a substantial development in person-computer interaction, presenting users with a more convenient and innovative means of communication [19]. Notably, speech recognition and text-to-speech technologies are employed by voice-enabled chatbots to aid smooth flow of communication and interaction among users [20]. Thus, this section explores two crucial areas of voice-enabled chatbot interfaces. Section 3.1 discusses the intricacies within the mechanisms of Natural Language Processing (NLP) algorithms which provides chatbots with the capability of understanding user queries and prompts within specific contexts and provide accurate responses. Section 3.2 on the other hand, explores the mechanism of spoken language conversion into accurate texts in the domain of speech recognition. It discusses the intricate process involved and the importance of improving convenience for users. Furthermore, it assesses the dynamics of speech recognition tools and discusses pertinent issues such as accuracy, conversion latency, etc.

## **2.4 Natural Language Processing**

Baby et al. [21] described NPL as the foundation for chatbots to function effectively, providing them with the capability to comprehend user queries and reproduce accurate responses contextually. The NPL technology is an integration of several algorithms that processes and analyzes human language [22]. This includes syntactic, semantic and sentiment analysis and processing [23]. A notable challenge commonly encountered in this process is the ability of the NPL to accurately determine user intentions and context of word use, requiring the incorporation of carefully written and sophisticated algorithms which can analyze and interpret these specificities of users input and queries [24]. The level of sophistication required was further explained by Worth [25], noting that NPL algorithms must have the capability to tackle language variations such as diction, abbreviations, slangs, etc., creating the necessity for word embedding to foster highly robust performance for serving a wide population and variety of users.

Addressing these challenges reveals the design of different approaches and NLP models for use in chatbot systems. One approach is the utilization of complex learning approaches such as the Recurrent Neural Networks (RNNs) [26] and the use of transformers which possess an increased level of superiority in the execution of vital functions such as generating certain language sequences and language modeling [27]. However, the project aims to utilise the transformer approach for the University of Dundee. As such, pre-trained language models such as OpenAI’s GPT (Generative pre-trained transformer stands out in this regard, significantly enhancing the capacity of PL-based chatbots to perform its designated functions [28]. These models have the capacity to produce logical and cohesive responses which are relevant to the context of use due to it’s massive training on a wide assemblage of texts, thus, most users are satisfied by the responses produced when queries are inputted [29]. As a result, by aligning these models with datasets on specific domains, programmers, and developers can enhance the ability of chatbots to comprehend queries and provide responses within specific domains, for instance, educational domains in this case [30].

Open AI’s GPT 3.5 model, while possessing exceptional capabilities as a language processor, a literature review on the advantages and limitations that accompany it is crucial. According to Caramancion [31], one major benefit of integrating GPT 3.5 into chatbots is the remarkable ability to produce contextually accurate outputs, with a high level of proficiency in terms of fluency and coherence. The model leverages a wide range of data texts which gives it the capability to understand complex language patterns amongst users and excel at comprehending semantics and syntax, facilitating seamless interactions with a wide range of users [32]. However, certain studies show that the huge reliance of GPT 3.5 on previous data presents crucial limitations. For instance, the AI may produce biased and inaccurate responses based on the inherent inaccuracies of the training data [33]. Furthermore, despite the range of applicability of GPT 3.5, optimal performance may not always be achieved and could vary depending on how complex queries might be [34]. For example, in some cases, the AI model might exhibit profound difficulty in comprehending nuanced queries or industry-specific questions, thus generating irrelevant responses. Thakkar [35] agreed, noting that in certain scenarios, the AI may prove difficult to understand as the responses might be complex to interpret, promoting inquiries behind the responses it generates and its decision-making process. Furthermore, ethical issues have been associated with the AI model and the dependence on cloud-based services and internet infrastructure could result in issues such as delay, latency and reduced reliability in performance [36], [37]. However, despite these challenges, GPT 3.5 boasts of a wide range of user-friendly capabilities that attract interest from various fields such as academia [36].

## **2.5 Voice-Enabled Chatbots**

A review on voice recognition is also crucial as it represents the foundation for the development of voice-enabled chatbot systems and interface [38]. Thus, this section examines literature on two crucial components of voice recognition: speech-to-text conversion and text-to-speech conversion. Speech-to-text conversion examines literature that discusses the processes through which spoken language is converted into text, exploring the specific mechanisms and potential obstacles in arriving at accurate transcriptions. Text-to-speech, on the other hand, analyses how natural sounds are synthesized from texts, also examining the mechanisms, technologies, and obstacles required for accurate translations into human-like voices.

### 2.5.1 Speech-to-text

Speech-to-text conversion, also referred to as automatic speech recognition (ASR), constitutes a vital part of voice-enabled chatbots, providing the capability to convert spoken words into texts [39]. Speech-to-text conversion usually involves multiple stages which include: analysis and processing of signals which involves the processing of raw audio signals in order to improve their quality for easy recognition and extract useful information therein [40]. Several other studies revealed a consensus in the process which involves audio filtering, normalization, and extraction with techniques such as the Mel-frequency Cepstral Coefficients (MFCCs) [41], [42], [43]. The next stage involves feature extraction which essentially fulfils the function of capturing important features of the audio signals. The purpose of this stage is to model the speech signals into a format that can be easily processed in later stages [44]. Examples of features utilized in this stage include filter banks and spectral features which have the ability to recognize inputted information [45]. More dated literature identified one of the next stages as the use of statistical models to identify the captured text features as either phonetic elements or sub-word elements [46], [47]. The critical role language modeling plays in speech-to-text recognition systems was underscored by different literature. For instance, Chou et al. [48] noted that language modelling integrates linguistic elements to improve the accuracy of speech recognition and boost fluency. In the same vein, Kushal [49] further noted that certain features such as statistical properties of languages – the English Language for instance, and utilise them to provide direction and guidance for the entire recognition process, upscaling the overall coherence of the text being transcribed from speech. Such statistical properties include word probabilities and sequences [50], [51].

A prominent example of an advanced speech-to-text AI is the Assembly AI which offers a highly sophisticated speech-to-text capability, prioritizing reduced latency and enhanced accuracy [52], making it a viable option for integration into the chatbot being designed for the University of Dundee. Critically examining the features is crucial in order to uncover nuanced benefits and potential limitations or disadvantages. One prominent feature of Assembly AI which stands out is its focus on increased speech-to-text conversion accuracy, which ensures that responses generated from user input are highly reliable and usable [53]. According to [54], this feature is essential for chatbots to properly contextualize responses based on the queries or inputs. However, Malage et al. [54] also noted that despite its advantages of high accuracy, a noisy environment could inhibit the accurate transcription of speeches. Additionally, dialectic difference might not be accurately represented in training data, hence, certain accents could be transcribed inaccurately. Additionally, despite the Assembly AI offering a highly robust platform system for converting speech to text, it may lack some advanced features which other speech recognition platforms such as Google Cloud Speech-to-text offer. However, it supports major languages such as English, Spanish, Italian, German, Portuguese and Dutch, making it highly appealing to a wide audience.

Hence, speech-to-text conversion, while a complex and sophisticated process, is a vital component of voice-enabled chatbot technologies, encompassing an array of procedures such as signal processing, extraction of voice features, and modeling.

### 2.5.2 Text-to-Speech

Another crucial component of voice-enabled chatbots is text-to-speech (TTS) AI technology which has the capacity to convert inputs in form of texts into speech. The AI technology functions by utilizing specially designed algorithms incorporated into it [55]. Literature from earlier and recent studies reveal different approaches to Text-to-speech synthesis which are still being utilized. For instance, Iida and Campbell [56] explained that one of the crucial and rudimentary approaches is concatenative synthesis which involves the combination of previously recorded speech fragments to synthesize progressive utterances (text conversion to speech). The process utilizes a database consisting of several speech elements such as phenomes in order to create speech through concatenation depending on the nature of the inputted text [56]. More recent studies such as [57] also underlined a similar process, explaining how concatenative synthesis is utilized to derive speech segments that bear high similitude to natural sounds. Another approach to text-to-speech is an approach referred to as parametric synthesis, involving the deployment of mathematical models to create speech waveforms modeled after certain acoustic features such as unique voice properties, and linguistic features such as diction [58]. Thus, the model is composed of two main parts – a module that analyzes text thoroughly and another module that synthesizes the analyzed text and generates speech in form of waves [59], [60].

However, massive advancements in language learning and deep learning have transformed TTS synthesis, culminating in the creation of network-based models which have the capacity to produce natural-sounding outputs [61]. The literature observed identified two main approaches. Firstly, the utilization of WaveNet which can synthesize speeches in form of waves by analyzing different voice samples [61]. Hence by thoroughly examining fine details of different sorts of complex speech, WaveNet synthesizes speech from texts and outputs them with a high level of similitude to natural voice. Secondly, transformer-based architectures, such as Tacotron have a tremendous capacity to synthesize text from speech by integrating text and acoustic properties [61], employing the use of certain mechanisms to synchronize texts with matching acoustic features, making it possible to ultimately generate natural-sounding speech.

A prominent example of WaveNet-based text-to-speech synthesizer is Google Cloud’s text-to-speech API, offering a reliable method for transforming text inputs into speech with high similitude to human voice [62]. Thus, it offers a highly useful means for improving interactions with AI Voice-Enabled chatbots in different systems. One primary advantage Google Cloud’s TTS API offers is the capacity to synthesize voices that sound natural by leveraging on sophisticated network-enabled models to create various forms of speeches that mimic human communication patterns [63]. Additionally, options for customization are offered by Google Cloud’s API, giving developers the option of creating tailored voice responses to align with the chatbots, thus fostering increased immersion and engagement among users [64]. While WaveNet-based TTS such as Google Cloud’s TTS API offers several advantages, limitations exist such as additional developmental and configuration requirements which increase the complexity of the entire TTS system [65].

## **2.6 User Experience and Accessibility**

A crucial consideration in the development of generative AI voice-enabled chatbots for organizational use is user experience and accessibility [66]. Thus, this section explores literature on the subject and highlights the different strategies which can be employed in order to ascertain the delivery of an engaging and interactive user experience. One of the foundations for the development of AI voice-enabled chatbots is a user-centered design which profoundly emphasizes a thorough understanding of user behavioral patterns and preferences which determine mod designs [67]. This usually involves comprehensive research to gain deep insights into communication systems and patterns in specific contexts [67]. Examples of such research methods include surveys, interviews, etc. The entire process gives developers the required knowledge and information to create and tailor chatbot properties and functionalities to match user needs more accurately [67]. Complementarily, persona development is imperative to the overall functionality of generative AI chatbots in terms of enhancing users’ experiences and promoting accessibility. The importance lies in its creation of virtual representations of the intended user population by understanding their unique features, wants, and objectives, thus functioning as a guide for developers [68]. Hence, by encapsulating the different preferences of users, persona development promotes more robust systems that emphasize certain features that seamlessly align with users’ end goals [68].

A further review of other literature revealed other strategies utilized to enhance user accessibility and experiences. For instance, across diverse studies, compliance with web accessibility standards was consistently noted as a crucial consideration for developers in order to ensure the technology remains widely accessible to various categories of individuals with a variety of abilities [69], [70]. Such include: navigation with input devices such as keyboard, color contrast and screen-reader-harmony. These are commonly incorporated into the chatbots to ensure a smooth experience for all types of users [70]. Furthermore, developers employ the use of multimodal interaction to address user preferences and accessibility requirements due to the chatbot's support of varying modes of interaction such as text input and graphical interface which enables users to utilize it through a modality of their own choice [71].

## **2.7 Literature Gap**

The literature review extensively discussed the developments, dynamics and paradigms that exist in the realm of AI-based chatbots. Although the diverse literature examined the huge strides in the development of the technology, certain gaps were observed which are crucial to the research. For instance, contextual understanding was not thoroughly discussed as observed in [27], [29], [30] which only scratched the surface and provided a superficial analysis. Further limitations are underscored by the overreliance of chatbots on recognizing patterns and generating results based on pre-defined sets of rules, limiting the capacity to immediately grasp meaning based on certain contexts [31], [63], [67]. Very few literature explored potential solutions and recommended strategies to circumvent the problem. Hence, the gap presents opportunities for further research in the advancement of Natural Language Processing and more sophisticated Machine Learning modalities which offer chatbots the capacity to generate contexts simply by analyzing user inputs and dialogues.

## **2.8 Summary**

Discussed extensively in this literature review is the complex landscape of communication technology, particularly the development of AI-driven chatbots to redefine and repurpose communication in diverse contexts, notably in educational settings. Comprehensively analyzed was the generative AI voice-enabled chatbots to address this issue and provide relevant responses to queries inputted by users, thus fostering improved interactions [20]. By exploring the use of generative AI in educational settings, the literature identified the vast potentialities offered by AI-driven chatbots to transform communication within university communities. Through the use of advanced language processing such as OpenAI’s GPT 3.5, the chatbot aims to understand human speech and inputs, and reproduce contextually accurate responses, thus boosting user engagement and enhancing communication paradigms [31], [32]. Additionally, the literature reviewed emphasized on the importance of utilizing advance speech-to-text AI such as the Assembly AI and sophisticated text-to-speech AI such as Google Cloud’s API [52], [53], [63], [64], [65]. The studies reviewed noted that both AI systems can significantly improve the chatbot's capability in terms of capturing users’ attention through engaging voice interaction techniques. Moreover, the review underscored the importance of user accessibility. Several methods for improving this include user-centered design principles, persona development, ensuring compliance with web accessibility standards, and utilizing multimodal interactions [67], [68], [69], [70], [71]. Importantly, crucial literature gaps were identified. For instance, the need for further studies in building more robust text-to-speech systems and their compatibility in educational settings was highlighted.

In conclusion, while the importance of developing an interactive and AI-enabled chatbot was outlined by the literature analyzed and reviewed, limitations which must be considered exist. Nonetheless, utilizing technology to enhance communication within educational communities represents a huge leap in transforming communication and engagement. Hence, through a holistic integration of a variety of advanced AI technologies, this literature review has provided a comprehensive analysis of relevant studies to aid in the research.

# **3.0 Specification**

The University of Dundee website holds a great deal of important information and offers several services to students, staff, and visitors on the website. Many students encounter numerous difficulties during their time in the university, and many of these difficulties can be solved with the resources and information offered by the university. An average student has to contact the enquire center in other to gain information that is easily accessible to them, while some students have no idea how to access these services and where to gain useful information to help in their times of difficulty.

The purpose of this project is to improve accessibility to information for all members of the University of Dundee community by implementing an intuitive and responsive platform for users to interact with, while facilitating seamless navigation of the university website and ensuring efficient information retrieval.

## **3.1 Requirements**

This chatbot was initiated to enhance user experience on the University of Dundee website, by focusing on accessibility, navigation, and information retrieval. The primary method of data collection for this project involves direct engagement with the end user. The end users could be prospective students, current students, staff, and visitors in general. This was achieved through observational studies aimed towards understanding the needs, preferences, and challenges faced when using the university website. Complementary to observations made, user interactions with the website were so analyzed in a natural setting without their explicit awareness of users to avoid behavioral changes. This approach allowed for the gathering of genuine data, highlighting areas of the website users sought information most, where they found challenges navigating, and the overall experience. This led to understanding that many users are unaware of the majority of services offered by the university as they resort to taking the initiative to find solutions independently when the solution they seek can be offered by the university.

The combination of observational studies and direct user feedback provides a comprehensive understanding of the current state of the website in terms of accessibility, which helped inform the development of this chatbot by highlighting areas where user engagement and satisfaction could be significantly improved. Furthermore, this approach to data collection ensures a user-centered development process, aligning the chatbot’s capabilities with the user's needs and preferences, ensuring the project goals are met effectively.

To secure the successful implementation of this chatbot, a series of functional and non-functional requirements must be met.

### 3.1.1 Functional Requirement

These are functions the chatbot must deliver to meet user requirements. These requirements include:

* Information Retrieval: The chatbot must be able to provide users with accurate information on university programs, services, and facilities.
* Accessibility Features: The chatbot must include features that make it accessible to users with different mental and physical impairments. These features include voice input and output and displayed responses in text, compatible with screen readers.
* Navigation Assistance: the chatbot should assist users in navigating the university website by redirecting users to relevant pages or resources based on their inquiries.

### 3.1.2 Non-Functional Requirements

These requirements outline the quality standards this chatbot must meet to ensure it delivers consistency, efficiency, and secures user satisfaction.

* Performance: The chatbot should respond to user queries promptly, to maintain engagement.
* Usability: The interface should be dynamic, responsive, intuitive and user-friendly to promote engagement, ease of use and adaptability.
* Reliability: The chatbot must provide consistent, accurate responses to user queries.

## **3.2 Project Plan**

The project plan outlines the approach taken to develop and implement this Generative AI Voice Enabled Chatbot on the University of Dundee website. It elaborates in detail on the methodology used and timeline of this project.

### 3.2.1 Project Time Line

This project began on the first week of February. The first week of this project was dedicated to research. It focused mainly on deciding what website or service the chatbot would implement and conducting research on generative AI, chatbots, text-to-speech, and speech-to-text.

The Second and third weeks were dedicated to planning, design, and requirement analysis. It involved deciding what programming language would be used, what generative AI service, speech-to-text API, and text-to-speech API would be used, obtaining the AI credentials i.e. Open AI, Assembly AI, and Google Cloud, API keys, what web framework would be used to develop the chatbot backend, and what framework would be used to design the user interface.

By the fourth week, the development phase had begun. This phase lasted for three weeks and it involved

* Open AI GPT 3.5 turbo integration for Natural Language Processing
* Implementation of Assembly AI for real-time transcription (speech-to-text conversion)
* Implementation of Google Cloud text-to-speech API for voice response
* Backend development with Fast API and Python
* Frontend development with HTML CSS and jQuery.
* Debugging and Testing

By the seventh week, the development phase had come to an end. This week was dedicated to evaluation, user testing, and report writing.

### 3.2.2 Justification of Methodology

For the development and implementation of this chatbot, the Agile Methodology was selected as a guiding framework for several reasons.

1. Flexibility and Adaptability: due to its collaborative nature and its ability to accommodate the evolving needs of this project, agile allows for flexibility in the development of this chatbot, which is essential as this project encountered various rapid changes in the technology used.
2. Iterative Development: The iterative approach of agile benefits this project by allowing the project to be broken down into manageable units, which enables the focus on high-priority features first while making continuous improvements based on regular testing and user feedback. This iterative approach ensures that the chatbot consistently aligns with the user's needs and expectations.
3. Quality Focus: Continuous integration and testing are a core part of agile. Integrating regular testing cycles throughout the development process can ensure high-quality outputs at every stage of the development process, which can lead to a reliable and effective chatbot.
4. User-Centered Design: By integrating regular user testing and feedback, the project can ensure the chatbot's functionality and design are truly user-centric, addressing user needs and enhancing their experience on the University of Dundee website.

# **4.0 Design**

The design phase was very crucial in setting the foundation for this chatbot. This section outlines the initial design choices including tools, frameworks, programming languages, and the reasons behind the decisions made. Due to the chatbot's dynamic nature, some design choices evolved to better meet the project goals during implementation.

## **4.1 Main Design Choices**

### 4.1.1 Programming Languages

Python is a high-level programming language known for its readability and clear syntax. Python was chosen for its high versatility, compatibility, and extensive support for various machine learning and natural language processing libraries like Spacy and TensorFlow. Python’s availability of documentation, wide community support, and resources, helped with easily overcoming and mitigating development hurdles, ensuring time management.

HTML (Hyper Text Markup Language) is a markup language for creating webpages. It creates a basic structure of the website which can be modified with other programming languages like CSS and JavaScript. The use of HTML ensures smooth Integration into the university website as it is highly compatible with web frameworks.

CSS (Cascading Style Sheets) is a stylesheet language used to describe the presentation for documents written in markup languages, It defines how elements are rendered on the website. CSS was chosen for its ability to define visual presentation of the website i.e. looks, layout etc., CSS3 was used in this project as it has various features for responsive design that adapts to different screen sizes.

JavaScript is a powerful high-level, versatile programming language used for web development, It adds interactive dynamic features to a webpage. jQuery which is a small, fast JavaScript library was chosen for this project for its efficiency in manipulating DOM (Document Object Model), event handling, simplicity, and animations. jQuery simplifies programming with JavaScript, making it easier to develop complex features with fewer lines of code.

### 4.1.2 Natural Language Processing with Open AI

Open AI is an artificial intelligence research laboratory which has made significant contributions to the AI field through research and development of top-notch technologies and tools. It houses some of the most notable advanced neural network models like the GPT (Generative Pre-trained Transformer) series, which includes the GPT 3, 3.5, and 4, which are capable of natural language processing and generating human-like text. The Open AI’s GPT 3.5 was adopted as the NLP core engine, for its advanced exceptional ability to generate human-like text, process and give contextual responses to a wide array of queries. In addition to its advanced abilities, the GPT model is easy to integrate and scalable.

### 4.1.3 Speech-to-Text with Assembly AI

Assembly AI is an artificial intelligence company that specializes in developing speech recognition and audio processing technologies. These advanced speech AI models are capable of transcribing, analyzing, and recognizing speech with high accuracy. These models support over 17 languages and are designed to be highly accurate with low latency. Although, Assembly AI was not the initial choice for this project, its advanced ability far exceeded the expectations and also the requirement for this chatbot. Despite the wonderful capabilities of Assembly AI speech models, it is also easy to integrate with low operational costs, hence why it was chosen for this project.

### 4.1.4 Text-to-Speech with Google Cloud

The Google Cloud Text-to-Speech API is one of the many services available on the Google Cloud platform. It is a powerful tool that allows for the conversion of text to natural-sounding speech, as it leverages Google's advanced deep-learning neural network algorithms to produce voice synthesis. It offers various speaking styles, emotional tones for different requirements while supporting a wide range of voices and languages. This API was chosen for its advanced capabilities and its straightforward integration with Python.

### 4.1.5 Backend Web Framework FastAPI

FastAPI is a modern, fast, high-performance web framework for building APIs, it is written with Python version 3.6 and above. FastAPI stands out for its fast performance thanks to its ability to handle concurrent users efficiently, and its asynchronous support. FastAPI was selected for this project above other web frameworks like Flask and Django for its superior speed, scalability, and easy integration in Python. This modern technology facilitates faster development cycles due to its automatic Swagger documentation feature.

## **4.2 Unified Modeling Language (UML)**

UML diagrams were developed to outline the architecture, software design, data flow and interaction of this chatbot, these designs provide a blueprint for the chatbot development process, system components and user interactions.

### 4.2.1 Use Case Diagram

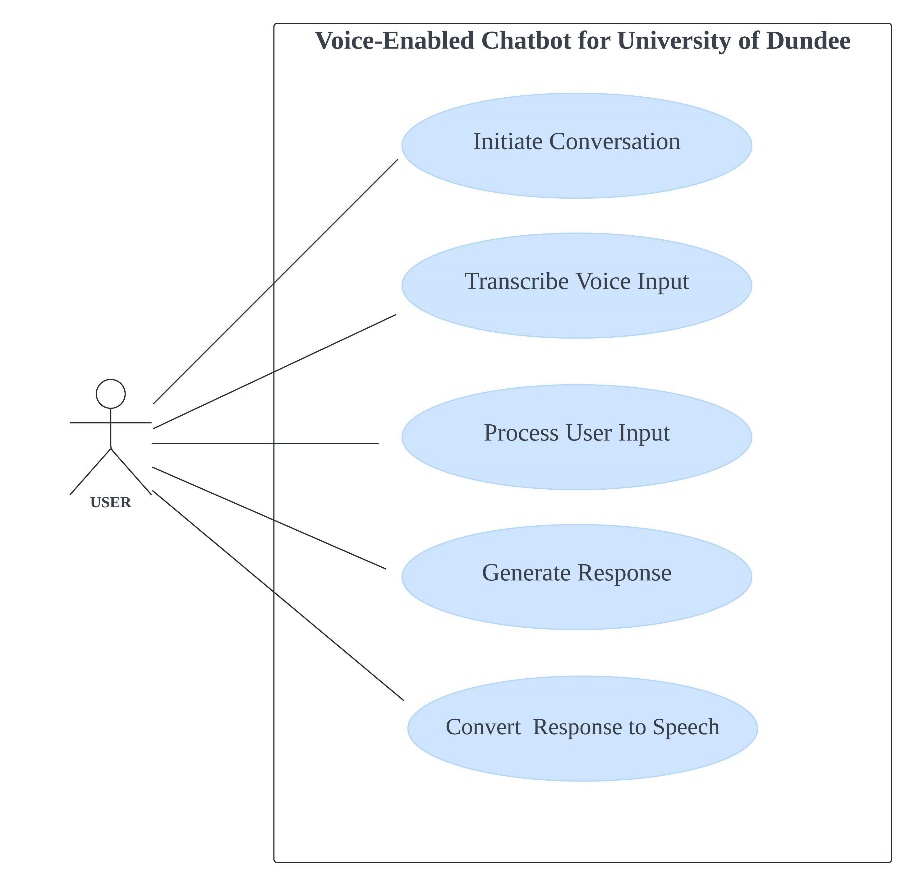
This diagram is a graphical representation of how users interact with the system. It helps in understanding what the chatbot does from a viewer's perspective, focusing on tasks and interactions rather than how the chatbot works internally.

Figure 4.1: Use Case Diagram

### 4.2.2 Sequence Diagram

This diagram shows how processes or objects interact; it displays the sequence of messages exchanged between different parts of the chatbot to carry out a specific process. It maps out interactions in chronological order, highlighting the order of operations.

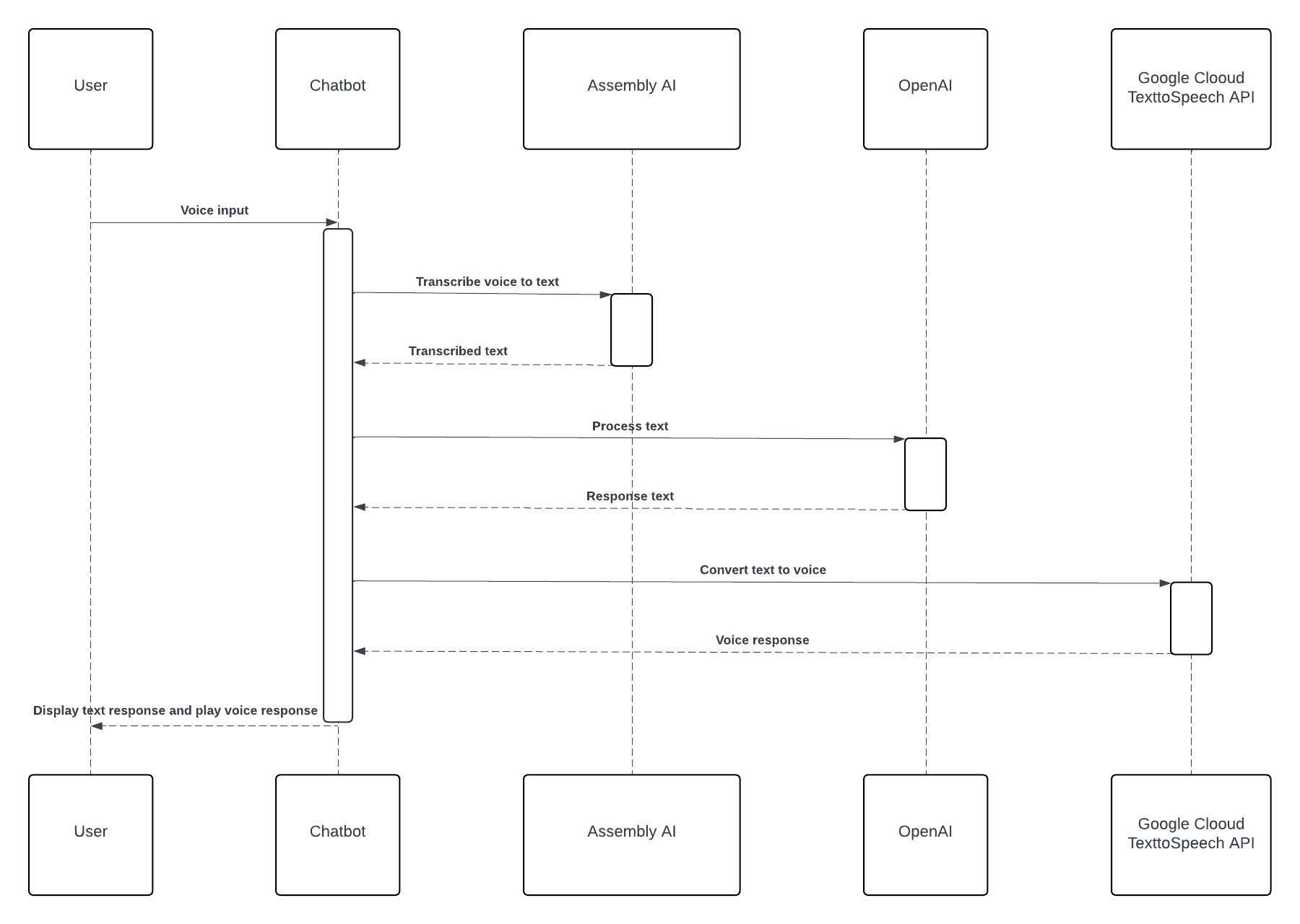


Figure 4.2: Sequence Diagram

### 4.2.3 component Diagram

Component diagrams are used to illustrate how the chatbot is divided into components, their relationships, arrangement, and dependencies, these components encapsulate the chatbot's functionality.

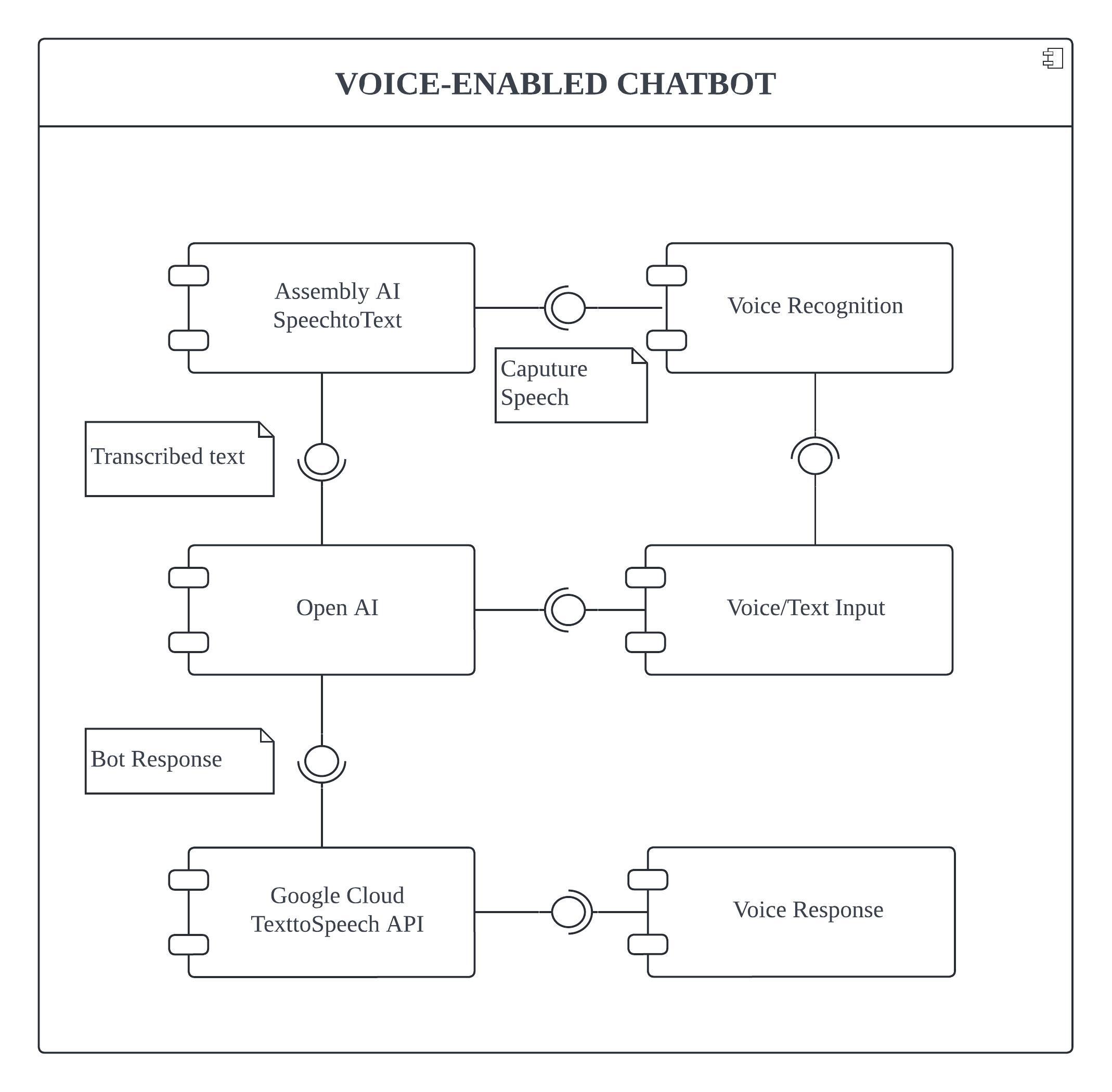


Figure 4.3: Component Diagram

# **5.0 Implementation**

The implementation of this chatbot was a complex process that spanned several stages, from initial design to final deployment. As discussed in section 3 this project adopts Agile methodology hence the use of an iterative approach for implementation. This project was organized into a series of phases, each phase dedicated to developing and refining key features of the chatbot based on predefined requirements.

## **5.1 System Architecture**

System Architecture is a high-level design that defines the structure and behavior of a system. It displays the relationships between system components and how they work together. Figure – shows the relationship between the different components of this chatbot, it displays the process taken from when the chatbot receives user input to when it plays/displays a response.

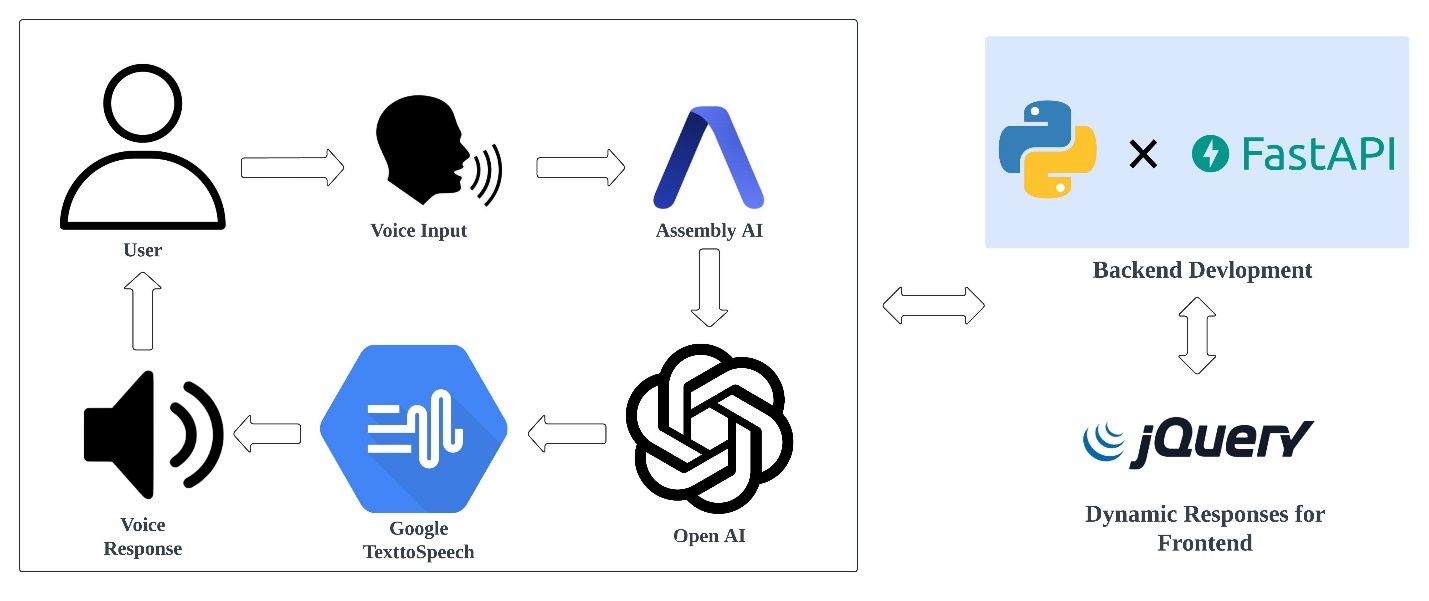


Figure 5.1: System Architecture

## **5.3 Phase-Based Development**

### 5.3.1 Phase 1: Open AI Integration

Given the nature of Open AI, integrating the GPT 3.5 model was effortless, there were no challenges faced during this phase. Before the integration, a Python virtual environment was created on Visual Studio Code IDE and the necessary packages were installed. Subsequently, an Open AI API key was generated from the Open AI website, this key was stored in a .env file. The .env file was then imported into the main.py file, using Python dotenv library, which ensures the API key is securely stored.

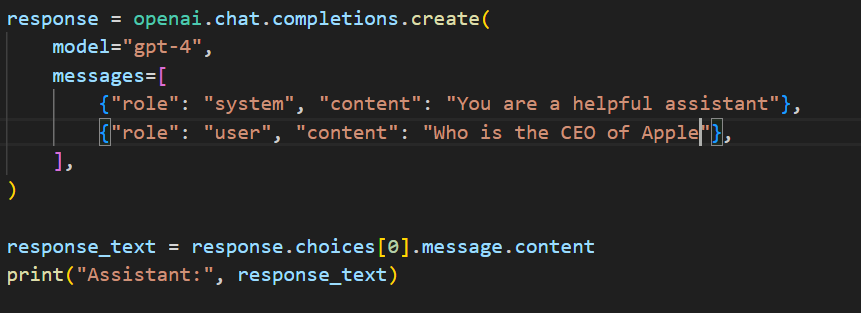
****

Figure 5.2: Code Snippet Open AI Integration

**5.3.2 Phase 2: Voice Input and Output Integration**

Analogous to Phase 1, the packages required to implement Assembly AI’s real-time transcription and Google Clouds text-to-speech API were installed on the virtual environment created in Phase 1. Both API keys were generated for their respective websites, and stored in the same .env file.

Implementing the voice feature for this chatbot was very challenging, this resulted in changing the speech-to-text and text-to-speech APIs multiple times. The initial AI model proposed was Open AI Whisper API which, in contrast to OpenAI's GPT model, proved ineffective. Efforts to implement the whisper-1 model were unsuccessful despite extensive attempts. In search of a solution, the discovery of Assembly AI and its remarkable properties emerged. Integrating Assembly AI’s real-time transcription feature was seamless due to its high compatibility with Python. The failure to utilize the Whisper API influenced the decision to adopt Google Cloud's Text-to-Speech API, given that whisper-1 was intended for both speech-to-text and text-to-speech conversion. Similar to Assembly the alternative was Eleven Labs AI for voice synthesis. Implementing Eleven Labs alongside the Pydub library was straightforward. However, midway through the integration phase, the Eleven Labs API became unresponsive despite all efforts to resolve the issue. This situation led to the exploration and discovery of Google Cloud's Text-to-Speech (TTS) API. Integrating Google Cloud TTS API proved to be effortless, as it did not require additional Python libraries to function, unlike the Whisper API and Eleven Labs AI.

### 5.3.3 Phase 3: Backend Development

At this phase, the task was to integrate Phase 1 and Phase 2 with the backend interface using FastAPI. Compared to other Python web frameworks such as Flask and Django, development with FastAPI is less complex, quicker, and more straightforward, requiring only a single main.py file and minimal code. To facilitate the reception of user input, which is subsequently forwarded to the GPT response engine, an endpoint to handle POST requests was created. This POST request is defined by `@app.post("/chatbot")`, configured specifically to receive the user's input.

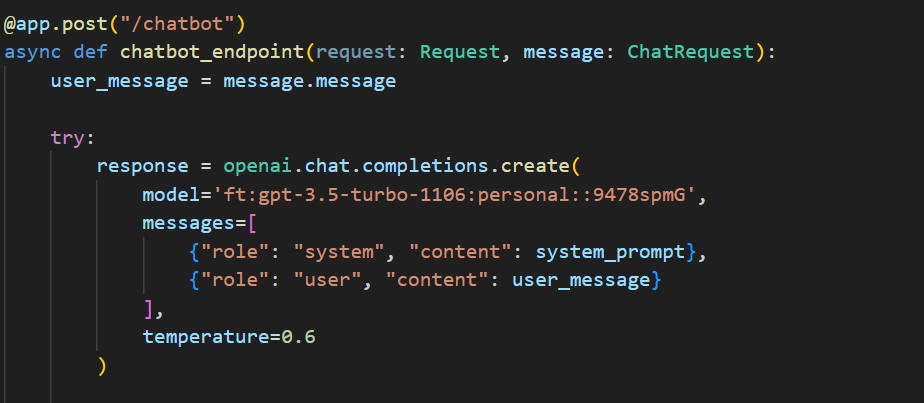


Figure 5.3: Code Snippet Chatbot Response Post Request

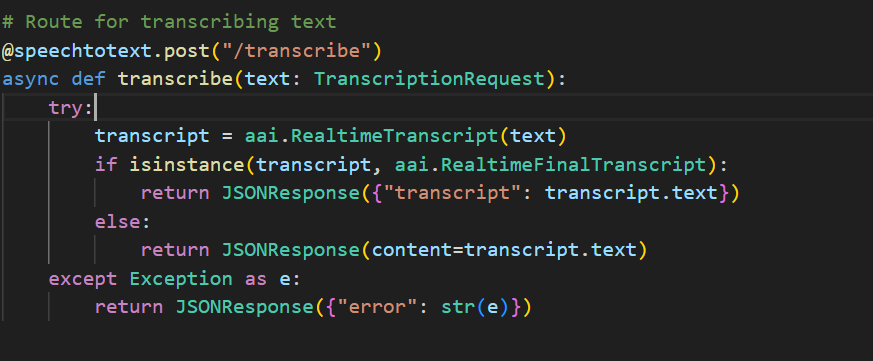
The asynchronous function “chatbot\_endpoint” processes the data sent by the user to the “/chatbot” endpoint. The data which is received as part of the request is extracted and sent to the Open AI’s API. The API response is captured and appended to the “chat\_log” for record-keeping. The same method is applied to capturing user's voice input; instead of receiving text the “/transcribe” POST endpoint is designed to receive speech and transcribe it to text using Assembly AI.

Figure 5.4: Code Snippet Assembly AI Transcription Post Request

The text-to-speech conversion process is initiated by a function named `synthesize\_text`, which accepts a string as input and utilizes the Google Cloud Text-to-Speech API to convert this string into speech. This function is configured to a specific voice model and language, and it outputs the audio in MP3 format. Subsequently, a new POST endpoint, `/tts`, is set up to receive a bot response and uses this function to convert the response into audio. However, for the audio to be accessed, it must be served, which is facilitated by a GET request. Utilizing the GET endpoint `@app.get`, a request is made to serve the previously stored MP3 audio file from the server, returning a file response that enables the audio file to be played. Finally, for FastAPI to receive any request, all POST requests must return a JSONResponse.

### 5.3.4 Phase 4: Frontend Development

This Phase includes utilizing HTML, CSS, and jQuery to develop a dynamic and responsive chat interface that supports both text and voice interactions. The key enhancements include real-time updates to the chat interface in response to user input and bot response, and allowing users to effortlessly switch between typing and speaking. This process was guided by a commitment to the requirements stated in section 3, with extensive testing to ensure compatibility. Mounting the static and HTML files to the FastAPI app was straightforward using Jinja2 template, and other FastAPI libraries.

### 5.3.5 Phase 4: Finetuning Job

This is the last implementation Phase, which involved an iterative process of training the GPT model to become a helpful assistant on the University of Dundee website. This process required a dataset containing a system prompt, user role (user query), and assistant role (bot response), which are questions likely asked by the user and assistant response. This information was gathered from the University of Dundee A-Z helpful guide for students, this dataset was stored in a JSONL file and the model was trained on the Open AI website by implementing a Python program, using the “gpt-3.5-turbo-1106” model.

Figure 5.5: Fine-Tune Job

# **6.0 Evaluation and Testing**

The evaluation and testing for the Generative AI Voice-Enabled chatbot was designed to assess its usability, functionality, accessibility and performance. The method used was qualitative research which involved collecting user feedback to better understand their experience and opinion on the chatbot, the data-gathering technique used was an interview. Participants were a group of post-graduate students from the University of Dundee, faculty of science and engineering. It was a predominantly male group with the age group of 25 and above. Participants were asked to engage with the chatbot through a series of tasks designed to explore the chatbot’s features, these tasks included utilizing the voice feature, using chatbot prompt to navigate the website, and retrieving information from the chatbot. After exploring the chatbot, participants were asked to engage in an interview consisting of 15 questions which was audio recorded, to collect their feedback. All participants were informed of the study’s purpose, and their consent was obtained prior to participation, anonymity and data protection measures were strictly adhered to, in compliance with the university’s ethics guidelines. Subsequent to conducting the interview, the audio was transcribed and the recording was destroyed, the information gathered underwent analysis, and these were the results.

## **6.1 Results**

The evaluation yielded important findings. Highlighting the chatbot's strengths, and areas of improvement.

* **Usability**: Feedback indicated the chatbot’s interfaces was user-friendly and easy to navigate. Participants appreciated the seamless transition between text and voice interactions and found the voice interaction better than typing.
* **Performance**: Participants appreciated the quality of the voice recognition feature, its high accuracy, and its ability to recognize words in different accents. Although participants were content with the performance and speed of the bot response, some were not satisfied with the accuracy, saying “You get the wrong answers without the right question” or “I didn’t like some of the results I got because they were out-of-date”.
* **Impact**: Participants believed that implementing this chatbot on the University of Dundee website would have a positive impact on our community. They stated that it would make access to information easier, it would reduce the time spent browsing the website for relevant information, and it would keep students and staff up-to-date with the latest information.

Overall participants were satisfied with most of the chatbot’s functionalities, and they stated, they would definitely recommend it to fellow students of the University of Dundee, they also indicated the chatbot could be used by individuals with certain impairments, due to the voice input feature. Although the evaluation provided valuable insights, there are still limitations, and areas to be improved. Some participants stated they would want the chatbot to link directly to the web pages with the required information, while others would like the chatbot to include up-to-date information on current events based on queries made. Future studies could include a broader demographic to highlight more problem areas, which could provide further insights into how to perfect the chatbot to best suit the University of Dundee website.

The comprehensive evaluation and testing of the Generative AI Voice-Enabled Chatbot demonstrates its value as a transformative tool for the University of Dundee website. By addressing identified problem areas and leveraging user feedback, the chatbot can evolve to meet the dynamic needs and preferences of the university’s community, by enhancing accessibility engagement and information retrieval across campus.

# **7.0 Description of the Final Product**

The final product of this project is a sophisticated, state-of-the-art Generative AI Voice-Enabled Chabot that can be seamlessly integrated into the University of Dundee website. Designed to enhance user experience for students, staff, and visitors on the website, the chatbot serves as a dynamic, responsive interface for accessing information, navigating the website, and engaging with the University’s community. Below is an overview of the specifics of its design, key functionalities and characteristics.

## **7.1 Interactive User Interface**

* Functional Design: The chatbot interface features a seamless blend with the University website, maintaining a minimalistic, clean, user-friendly interface that invites interaction without overwhelming users with unnecessary complexity.
* Adaptive Interaction Modes: By recognizing the diverse preferences of its user base, the chatbot offers dual modes of interaction, text and voice offering flexibility and accessibility, responses are also provided in text and speech format.

## **7.2 Advanced Core Functionalities**

* Information Retrieval: The chatbot excels in providing and fetching information across a broad spectrum including, course details, campus facilities, administrative programs and much more. It leverages Open AI’s GPT-3.5 model which has been fine-tuned to understand and respond to queries relating to the university, with high degree of accuracy and contextual relevance.
* Navigation Assistance: Beyond information retrieval, the chatbot guides users through the website, by recommending relevant webpages, resources, or services based on inquiries made. This feature is particularly useful for visitors and new students who are unfamiliar with the university landscape.
* State-of-the-art Voice Interaction: Integrating Assembly AI’s and Google Cloud’s text-to-speech APIs, Users can easily switch from text to voice, enjoying a high-quality voice interaction experience which includes reliable speech recognition and natural-sounding responses.

## **7.3 Accessibility and Performance**

The chatbot is designed to be fully accessible, by featuring text-to-speech for visually impaired users, easy keyboard navigation and easy-to-understand responses suitable for all users. Being built on FastAPI, the chatbot’s backend is optimized for fast response, browser compatibility and high performance even during peak of traffic. This ensures instantaneous response to user queries. Cloud-based technologies, provide computational efficiency keeping operational costs to a minimum while delivering consistent, reliable service. As a result of the Fine-Tuning job, the chatbot uses machine learning to learn from previous interactions, gradually improving responses and recommendations to offer a more accurate and reliable experience over time.

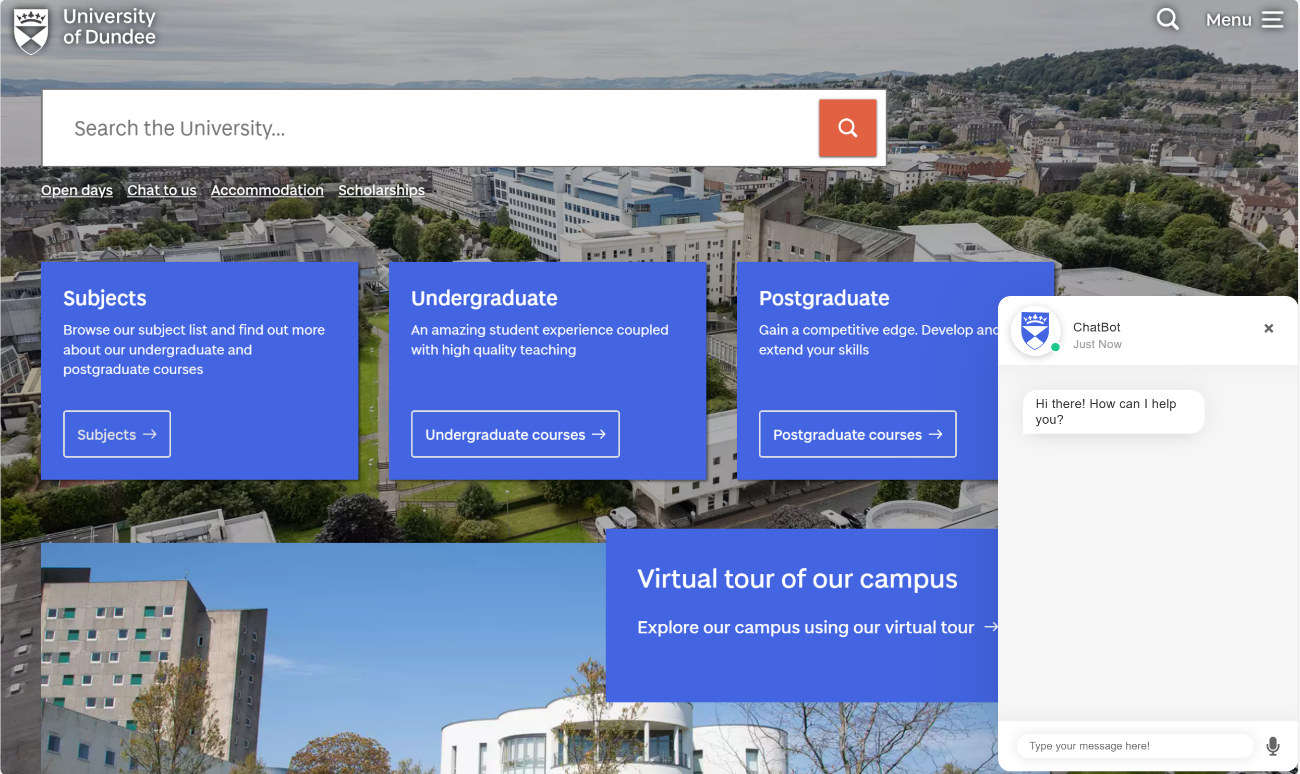


Figure 7.1: Final Product

# **8.0 Appraisal**

This project has been a significant undertaking that utilizes innovative technologies to enhance the digital experience of the University’s community. The decision to utilize these technologies is essential as they ensure high accuracy, reliability, and a seamless user experience. For future projects, minimizing the AI technologies used will help to reduce operational costs, rather than paying for multiple, only one payment will be required. FastAPI’s performance and effortless integration, were crucial to this project's success, retaining this choice for a similar project would be highly recommended.

The phase-based approach allowed for continuous improvement throughout this project. However, incorporating frequent user testing sessions at early stages would help identify usability issues sooner, allowing for quicker adjustment, also implementing a more flexible timeline in the project plan would help prevent the impact of challenges in future.

Although jQuery offers more features for less code, the use of React JavaScript would be highly recommended to prevent challenges with mounting static files in the FastAPI application, this will also result in a far more dynamic user interface if combined with NEXT JS.

Overall, the experience gained from developing this Generative AI Voice-Enabled chatbot will undoubtedly improve and inform future projects, driving further development of innovative digital technologies in the University of Dundee community.

## **8.1 Legal, Social, Ethical and Professional Issues**

When deploying the Generative AI Voice-Enabled Chatbot, it is crucial to clearly address legal, social, ethical, and professional issues. These aspects play vital roles in ensuring the project aligns with societal standards and norms, promoting trust among users.

### 8.1.1 Legal Issues

* **Data Protection and Privacy**: In compliance with the General Data Protection Regulation GPDR and other relevant privacy laws, the chatbot must be designed to ensure the protection of user's data and user privacy, i.e. if any data is collected. It involves obtaining consent for any data collected, ensuring transparency about how the data will be used, and implementing security measures to protect personal information.
* **Property Rights**: The development of AI or machine learning technologies involves the use of datasets that may be protected by copyrights. It is important to secure and acknowledge the source, the necessary license, or use open-source data to avoid copyright infringement.
* **Accessibility Compliance**: The chatbot must comply with legal standards for digital accessibility, such as the Web Content Accessibility Guidelines WCAG, to guarantee accessibility to all users including disabled users, aligning with legal requirements for inclusive digital environments.

### 8.1.2 Social Issues

* **Digital Inclusion**: The chatbot should be accessible and usable by all, including individuals with various degrees of digital literacy, to avoid digital divide. This ensures inclusivity to those who might not be comfortable with new technologies and provides multiple interaction modes to accommodate diverse preferences.
* **Impact on Employment**: The development of this chatbot might raise concerns about the impact on staff employment. It is important to view the chatbot as a tool to increase staff abilities, not replace them. It assists staff by attending to less tedious queries quickly and leaving more complex queries and tasks to them.
* **User Engagement**: The chatbot aims to provide a sense of community among users, it contributes positively to the university’s social environment encouraging active participation by providing a responsive and dynamic platform for interactions.

### 8.1.3 Ethical Issues

* **Bias and Fairness**: AI models can inherit biases from their training data, which can lead to unfair or prejudiced outcomes. Ethically, it is important to audit and update training datasets, to ensure the data is diverse and inclusive, mitigating discriminatory or biased interactions.
* **Transparency and Accountability**: Ethically, it is important for users to be aware they are interacting with an AI model. The chatbot should include a mechanism for feedback and reporting inaccurate or inappropriate responses to ensure the system is accountable.

### 8.1.4 Professional Issues

* **Quality and Reliability**: Devotion to professional standards during development ensures the chatbot is reliable, of high performance, accurate, and delivers fault tolerance. Regular maintenance and latest updates are essential for sustaining qualities and performance levels.
* Continuous Improvement: The AI field is continuously evolving. A commitment to improvements based on user feedback, technological advancement, and the evolving needs of the university, is crucial for maintaining relevance and value.

# **9.0 Summary and Conclusion**

The deployment of this Generative AI Voice-Enabled Chatbot represents an improvement in the communication dynamics within the educational sector. It also represents a forward-thinking approach toward enhancing user experience and information accessibility on the University of Dundee website. By utilizing advanced technologies, the chatbot provides an intuitive and seamless interface for users to easily access information, navigate the website and engage with the university’s online resources.

This report outlined in detail the comprehensive journey of the development process from early design to implementation and testing, guided by an Agile methodology.

Valuable lessons learned during the course of this project include the invaluable importance of flexible project planning and management, early and continuous testing, and the benefits of making user-centred design choices.

In conclusion, this project demonstrates how AI can transform digital platforms into a more interactive user-friendly environment which users can find extremely beneficial. The outcome not only reflects the success in achieving set goals but also a roadmap for future inventions in digital user experience. It also underscores the university's commitment to innovation, inclusivity, and providing an exceptional online experience for all users.

# **10 Future Work**

For future work, the first recommendation would be multilingual support. The University of Dundee houses students and staff from every corner of the globe. Expanding the chatbot to include multilingual support would make it accessible to a wider audience, reflecting the diverse linguistic background of the University’s community.

Secondly, implementing the chatbot on mydundee. Not all students or staff will visit the University of Dundee website, integrating with myDundee will make for easier access to all. Also, it could include more advanced personalized features such as learning user's preferences to outcome more tailored responses and recommendations.

Thirdly, Implementing a chat database to house previous chats. This chatbot does not have a database which results in the inability to ask follow-up questions, implementing a database will help the chatbot be conversational, enhancing user experience.

The fourth recommendation would be incorporating a geolocation feature to provide real-time campus navigation assistance for users, it could include mapping of university buildings which would be beneficial to new students.

Lastly, incorporating a wake word/hot word. Expanding the chatbot functionality to respond without contact, would broaden the diversity of audience of which the chatbot caters.

# **11 References**

1. M. Yamin, “Information technologies of 21st century and their impact on the society,” *International Journal of Information Technology*, vol. 11, pp. 759–766, August, 2019, doi: <https://doi.org/10.1007/s41870-019-00355-1>.
2. A. Haleem, M. Javaid, M. A. Qadri, and R. Suman, “Understanding the Role of Digital Technologies in education: a Review,” *Sustainable Operations and Computers*, vol. 3, no. 3, pp. 275–285, 2022, doi: <https://doi.org/10.1016/j.susoc.2022.05.004>.
3. J. C. Torrado, C. Bakke, and E. Gabarrón, “Chatbots and Children with Special Educational Needs Interaction,” *Lecture Notes in Computer Science*, pp. 443–452, January, 2023, doi: <https://doi.org/10.1007/978-3-031-35897-5_32>.
4. M. Mateos-Sanchez, A. C. Melo, L. S. Blanco, and A. M. F. García, “Chatbot, as Educational and Inclusive Tool for People with Intellectual Disabilities,” *Sustainability*, vol. 14, no. 3, p. 1520, January, 2022, doi: <https://doi.org/10.3390/su14031520>.
5. ‌M. Milana, U. Brandi, S. Hodge, and Tetyana Hoggan-Kloubert, “Artificial intelligence (AI), conversational agents, and generative AI: implications for adult education practice and research,” *International Journal of Lifelong Education*, vol. 43, no. 1, pp. 1–7, January, 2024, doi: <https://doi.org/10.1080/02601370.2024.2310448>.
6. G. Ilieva, T. Yankova, S. Klisarova-Belcheva, A. Dimitrov, M. Bratkov, and D. Angelov, “Effects of Generative Chatbots in Higher Education,” *Information*, vol. 14, no. 9, p. 492, September, 2023, doi: <https://doi.org/10.3390/info14090492>.
7. M. A. Kuhail, N. Alturki, S. Alramlawi, and K. Alhejori, “Interacting with educational chatbots: A systematic review,” *Education and Information Technologies*, vol. 28, no. 1, July, 2022, doi: <https://doi.org/10.1007/s10639-022-11177-3>.
8. C. Chakraborty, S. Pal, M. Bhattacharya, Sukanta Kumar Dash, and S. Lee, “Overview of Chatbots with special emphasis on artificial intelligence-enabled ChatGPT in medical science,” *Frontiers in artificial intelligence*, vol. 6, October, 2023, doi: <https://doi.org/10.3389/frai.2023.1237704>.
9. ‌L. Labadze, M. Grigolia, and L. Machaidze, “Role of AI chatbots in education: systematic literature review,” *International Journal of Educational Technology in Higher Education*, vol. 20, no. 1, October, 2023, doi: <https://doi.org/10.1186/s41239-023-00426-1>.
10. Aşkın Doygunel and Fatma Köprülü, “A study of the roles of school administrators in increasing the quality of school life through social responsibility projects in primary schools,” *Frontiers in Psychology*, vol. 13, September, 2022, doi: <https://doi.org/10.3389/fpsyg.2022.969638>.
11. ‌I. K. F. Haugeland, A. Følstad, C. Taylor, and C. Alexander, “Understanding the user experience of customer service chatbots: An experimental study of chatbot interaction design,” *International Journal of Human-Computer Studies*, vol. 161, p. 102788, May 2022, doi: <https://doi.org/10.1016/j.ijhcs.2022.102788>.
12. ‌Okan Yeti̇şensoy and Hıdır Karaduman, “The effect of AI-powered chatbots in social studies education,” *Education and Information Technologies*, February, 2024, doi: <https://doi.org/10.1007/s10639-024-12485-6>.
13. ‌S. Panda and R. Chakravarty, “Adapting intelligent information services in libraries: a case of smart AI chatbots,” *Library Hi Tech News*, vol. 39, no. 1, pp. 12–15, January, 2022, doi: <https://doi.org/10.1108/lhtn-11-2021-0081>.
14. ‌P. M. Mah, I. Skalna, and J. Muzam, “Natural Language Processing and Artificial Intelligence for Enterprise Management in the Era of Industry 4.0,” *Applied Sciences*, vol. 12, no. 18, p. 9207, September, 2022, doi: <https://doi.org/10.3390/app12189207>.
15. ‌G. Caldarini, S. Jaf, and K. McGarry, “A Literature Survey of Recent Advances in Chatbots,” *Information*, vol. 13, no. 1, p. 41, January, 2022, doi: <https://doi.org/10.3390/info13010041>.
16. ‌P. Singh and Rajalakshmi Krishnamurthi, “Chatbot-Based Android Application towards Security Using FCM,” *Lecture notes in networks and systems*, pp. 485–498, January, 2024, doi: <https://doi.org/10.1007/978-981-99-9811-1_39>.
17. ‌L. Gkinko and A. Elbanna, “The appropriation of conversational AI in the workplace: A taxonomy of AI chatbot users,” *International Journal of Information Management*, vol. 69, p. 102568, August, 2022, doi: <https://doi.org/10.1016/j.ijinfomgt.2022.102568>.
18. ‌Nataliya Mutovkina, “Chatbots as a Tool to Optimize the Educational Process,” *Advances in intelligent systems and computing*, pp. 15–25, January, 2021, doi: <https://doi.org/10.1007/978-3-030-80531-9_2>.
19. A.Banerjee, “Text and voice-enabled chatbot enhancing the user experience in career counselling domain,” esource.dbs.ie, 2019. <https://esource.dbs.ie/handle/10788/3712>
20. A. Jain and P. Bhati, “Comparative Analysis and Development of Voice-based Chatbot System for Differently-abled,” *Journal of Physics: Conference Series,* vol. 2273, no. 1, p. 012003, May 2022, doi: <https://doi.org/10.1088/1742-6596/2273/1/012003>.
21. ‌C. J. Baby, F. A. Khan, and J. N. Swathi, “Home automation using IoT and a chatbot using natural language processing,” *2017 Innovations in Power and Advanced Computing Technologies (i-PACT),* April, 2017, doi: <https://doi.org/10.1109/ipact.2017.8245185>.
22. ‌Y. Wang and C. Zhang, “Using the full-text content of academic articles to identify and evaluate algorithm entities in the domain of natural language processing,” *Journal of Informetrics*, vol. 14, no. 4, p. 101091, November, 2020, doi: <https://doi.org/10.1016/j.joi.2020.101091>.
23. V. K. and D. Gupta, “Unmasking text plagiarism using syntactic-semantic based natural language processing techniques: Comparisons, analysis and challenges,” *Information Processing & Management*, vol. 54, no. 3, pp. 408–432, May 2018, doi: <https://doi.org/10.1016/j.ipm.2018.01.008>.
24. ‌T.-Y. Chang and Y.-N. Chen, “What Does This Word Mean? Explaining Contextualized Embeddings with Natural Language Definition,” January, 2019, doi: <https://doi.org/10.18653/v1/d19-1627>.
25. ‌P. J. Worth, “Word Embeddings and Semantic Spaces in Natural Language Processing,” *International Journal of Intelligence Science*, vol. 13, no. 01, pp. 1–21, 2023, doi: <https://doi.org/10.4236/ijis.2023.131001>.
26. ‌K. M.Tarwani and S. Edem, “Survey on Recurrent Neural Network in Natural Language Processing,” *International Journal of Engineering Trends and Technology*, vol. 48, no. 6, pp. 301–304, June,2017, doi: https://doi.org/10.14445/22315381/ijett-v48p253.
27. ‌K. Subramanyam Katikapalli, R. Ajit, and S. Sivanesan, “AMMUS: A Survey of Transformer-based Pretrained Models in Natural Language Processing,” *Computation and Language*, August, 2021, doi: <https://doi.org/10.48550/arXiv.2108.05542>.
28. ‌T. N. Fitria, “Artificial intelligence (AI) technology in OpenAI ChatGPT application: A review of ChatGPT in writing English essay,” ELT Forum: *Journal of English Language Teaching*, vol. 12, no. 1, pp. 44–58, March,2023, doi: <https://doi.org/10.15294/elt.v12i1.64069>.
29. ‌A. Jojic, Z. Wang, and N. Jojic, “GPT is becoming a Turing machine: Here are some ways to program it,” *Computation and Language*, March, 2023, doi: <https://doi.org/10.48550/arxiv.2303.14310>.
30. L. Villa, D. Carneros-Prado, Adrián Sánchez-Miguel, C. C. Dobrescu, and Ramón Hervás, “Conversational Agent Development Through Large Language Models: Approach with GPT,” *Lecture notes in networks and systems*, pp. 286–297, January, 2023, doi: <https://doi.org/10.1007/978-3-031-48306-6_29>.
31. ‌K. M. Caramancion, “News Verifiers Showdown: A Comparative Performance Evaluation of ChatGPT 3.5, ChatGPT 4.0, Bing AI, and Bard in News Fact-Checking,” June, 2023, doi: <https://doi.org/10.48550/arxiv.2306.17176>.
32. ‌[1]K. I. Roumeliotis and N. D. Tselikas, “ChatGPT and Open-AI Models: A Preliminary Review,” *Future Internet*, vol. 15, no. 6, p. 192, June,2023, doi: https://doi.org/10.3390/fi15060192.
33. ‌K. Abramski, S. Citraro, L. Lombardi, G. Rossetti, and M. Stella, “Cognitive Network Science Reveals Bias in GPT-3, GPT-3.5 Turbo, and GPT-4 Mirroring Math Anxiety in High-School Students,” *Big data and cognitive computing*, vol. 7, no. 3, pp. 124–124, June, 2023, doi: <https://doi.org/10.3390/bdcc7030124>.
34. ‌Y. Kaneda et al., “Assessing the Performance of GPT-3.5 and GPT-4 on the 2023 Japanese Nursing Examination,” *Cureus*, August, 2023, doi: <https://doi.org/10.7759/cureus.42924>.
35. ‌K. Y. Thakkar and N. Jagdishbhai, “Exploring the capabilities and limitations of GPT and Chat GPT in natural language processing,” *Journal of Management Research and Analysis*, vol. 10, no. 1, pp. 18–20, April, 2023, doi: <https://doi.org/10.18231/j.jmra.2023.004>.
36. P. P. Ray, “ChatGPT: a Comprehensive Review on background, applications, Key challenges, bias, ethics, Limitations and Future Scope,” *Internet of Things and Cyber-Physical Systems*, vol. 3, no. 1, pp. 121–154, April, 2023, doi: <https://doi.org/10.1016/j.iotcps.2023.04.003>.
37. ‌G. Wang et al., “Potential and Limitations of ChatGPT 3.5 and 4.0 as a Source of COVID-19 Information: Comprehensive Comparative Analysis of Generative and Authoritative Information,” *Journal of Medical Internet Research*, vol. 25, no. 1, p. e49771, December, 2023, doi: <https://doi.org/10.2196/49771>.
38. ‌C. Bălan, “Chatbots and Voice Assistants: Digital Transformers of the Company–Customer Interface—A Systematic Review of the Business Research Literature,” *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 18, no. 2, pp. 995–1019, May 2023, doi: <https://doi.org/10.3390/jtaer18020051>.
39. ‌M. Malik, M. K. Malik, K. Mehmood, and I. Makhdoom, “Automatic speech recognition: a survey,” *Multimedia Tools and Applications*, November, 2020, doi: <https://doi.org/10.1007/s11042-020-10073-7>.
40. ‌V. A and D. V. Jose, “Speech to text conversion and summarization for effective understanding and documentation,” *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9, no. 5, pp. 3642–3648, October, 2019, doi: <https://doi.org/10.11591/ijece.v9i5.pp3642-3648>.
41. ‌A. B. Agrawal and R. Bhatia, “Development and Evaluation of a Voice Command App for Smartphone Interaction Using the Speech-to-Text API,” *Lecture notes in networks and systems*, pp. 247–256, January, 2024, doi: <https://doi.org/10.1007/978-981-99-8346-9_21>.
42. ‌A. U. Nasib, H. Kabir, R. Ahmed, and J. Uddin, “*A Real Time Speech to Text Conversion Technique for Bengali Language*,” (February, 01, 2018). Available: <https://ieeexplore.ieee.org/document/8465680>.
43. ‌Y. H. Ghadage and S. D. Shelke, “*Speech to text conversion for multilingual languages,”* (April, 01, 2016). Available: <https://ieeexplore.ieee.org/document/7754130/>
44. ‌S. Venkatasubramanian and R. Mohankumar, “A Deep Convolutional Neural Network-Based Speech-to-Text Conversion for Multilingual Languages,” *Advances in intelligent systems and computing*, pp. 617–633, January, 2022, doi: <https://doi.org/10.1007/978-981-16-9573-5_44>.
45. ‌M. R. Devi and T. Ravichandran, “A novel approach for speech feature extraction by Cubic-Log compression in MFCC,” *2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering*, pp. 182-186, February, 2013, doi: <https://doi.org/10.1109/icprime.2013.6496469>.
46. ‌W.-J. Wang, Y.-F. Liao, and S.-H. Chen, “RNN-based prosodic modeling for mandarin speech and its application to speech-to-text conversion,” *Speech Communication*, vol. 36, no. 3–4, pp. 247–265, March,2002, doi: <https://doi.org/10.1016/s0167-6393(01)00006-1>.
47. ‌L. Gu, Y. Gao, F.-H. . Liu, and M. Picheny, “Concept-Based Speech-to-Speech Translation Using Maximum Entropy Models for Statistical Natural Concept Generation,” *IEEE Transactions on Audio, Speech and Language Processing,* vol. 14, no. 2, pp. 377–392, March,2006, doi: <https://doi.org/10.1109/tsa.2005.860769>.
48. ‌J.-C. Chou et al., “Toward Joint Language Modeling for Speech Units and Text,” *Computation and Language*, October, 2023, doi: <https://doi.org/10.48550/arxiv.2310.08715>.
49. ‌Kushal Lakhotia et al., “On Generative Spoken Language Modeling from Raw Audio,” *DOAJ (DOAJ: Directory of Open Access Journals)*, January, 2021, doi: <https://doi.org/10.1162/tacl_a_00430>.
50. ‌C. Wei, Y.-C. Wang, B. Wang, and C.-C. . J. Kuo, “An Overview on Language Models: Recent Developments and Outlook,” *Computation and Language*, July, 03, 2023. <https://arxiv.org/abs/2303.05759> (accessed August, 13, 2023).
51. T. Tanaka, R. Masumura, T. Moriya, and Y. Aono, “Neural Speech-to-Text Language Models for Rescoring Hypotheses of DNN-HMM Hybrid Automatic Speech Recognition Systems,” *IEEE Xplore*, November, 01, 2018. <https://ieeexplore.ieee.org/abstract/document/8659622> (accessed March,31, 2023).
52. ‌R. V. Malage, H. Ashish, Sanjana Hukkeri, E. Kavya, and R. Jayashree, “Low Resource Speech-to-Speech Translation of English videos to Kannada with Lip-Synchronization,” *2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 1680-1687,* May 2023, doi: <https://doi.org/10.1109/iciccs56967.2023.10142578>.
53. Assembly AI (n.d.). “*Speech AI use cases for Learning Management Systems,” News, Tutorials, AI Research* [online]. Available: <https://www.assemblyai.com/blog/speech-ai-use-cases-learning-management-systems/> (accessed March, 17, 2024).
54. ‌R. Zhang and D. Zou, “Types, purposes, and effectiveness of state-of-the-art technologies for second and foreign language learning,” *Computer Assisted Language Learning*, vol. 35, no. 4, pp. 1–47, April, 2020, doi: <https://doi.org/10.1080/09588221.2020.1744666>
55. Z. Leini and S. Xiaolei, “Study on Speech Recognition Method of Artificial Intelligence Deep Learning,” *Journal of Physics: Conference Series*, vol. 1754, no. 1, p. 012183, February, 2021, doi: <https://doi.org/10.1088/1742-6596/1754/1/012183>.
56. [1]A. Iida and N. Campbell, “Speech Database Design for a Concatenative Text-to-Speech Synthesis System for Individuals with Communication Disorders,” *International Journal of Speech Technology,* vol. 6, no. 4, pp. 379–392, October, 2003, doi: <https://doi.org/10.1023/a:1025761017833>.
57. ‌S. P. Panda and A. K. Nayak, “A waveform concatenation technique for text-to-speech synthesis,” *International Journal of Speech Technology*, vol. 20, no. 4, pp. 959–976, October, 2017, doi: <https://doi.org/10.1007/s10772-017-9463-8>.
58. ‌S. Takamichi, T. Toda, Y. Shiga, S. Sakti, G. Neubig, and S. Nakamura, “Parameter Generation Methods With Rich Context Models for High-Quality and Flexible Text-To-Speech Synthesis,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 2, pp. 239–250, April, 2014, doi: <https://doi.org/10.1109/jstsp.2013.2288599>.
59. S. Suzić, T. V. Delić, S. Ostrogonac, S. Đurić, and D. J. Pekar, “Style-Code Method for Multi-Style Parametric Text-to-Speech Synthesis,” *SPIIRAS Proceedings,* vol. 5, no. 60, p. 216, October, 2018, doi: <https://doi.org/10.15622/sp.60.8>.
60. ‌Jürgen Trouvain and Bernd Möbius, “Speech Synthesis: Text-To-Speech Conversion and Artificial Voices,” *Springer eBooks*, pp. 3837–3851, October, 2019, doi: https://doi.org/10.1007/978-3-030-02438-3\_168.
61. ‌N. Kaur and P. Singh, “Conventional and contemporary approaches used in text to speech synthesis: a review,” Artificial Intelligence Review, November, 2022, doi: <https://doi.org/10.1007/s10462-022-10315-0>.
62. ‌R. Reimao and V. Tzerpos, “FoR: A Dataset for Synthetic Speech Detection,” *2019 International Conference on Speech Technology and Human-Computer Dialogue (SpeD),* October, 2019, doi: <https://doi.org/10.1109/sped.2019.8906599>.
63. ‌D. Chauhan, C. Singh, R. Rawat, and M. Dhawan, “Evaluating the Performance of Conversational AI Tools,” *Conversational Artificial Intelligence,* pp. 385–409, January, 2024, doi: <https://doi.org/10.1002/9781394200801.ch24>.
64. ‌T. Adamson, C. Burton Lyng-Olsen, K. Umstattd, and Marynel Vázquez, “Designing Social Interactions with a Humorous Robot Photographer,” *HRI ’20: Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot* Interaction March, pp. 233-241 ,2020, doi: <https://doi.org/10.1145/3319502.3374809>.
65. ‌L. Boonstra, “Implementing a Dialogflow Voice Agent in Your Website or App Using the SDK,” *Apress eBooks*, pp. 301–327, January, 2021, doi: <https://doi.org/10.1007/978-1-4842-7014-1_12>.
66. A. M. Al-Abdullatif, “Modeling Students’ Perceptions of Chatbots in Learning: Integrating Technology Acceptance with the Value-Based Adoption Model,” *Education Sciences*, vol. 13, no. 11, p. 1151, November, 2023, doi: <https://doi.org/10.3390/educsci13111151>.
67. Y. Kim, S. Myeong, and M. J. Ahn, “Living Labs for AI-Enabled Public Services: Functional Determinants, User Satisfaction, and Continued Use,” *Sustainability*, vol. 15, no. 11, p. 8672, January, 2023, doi: <https://doi.org/10.3390/su15118672>.
68. ‌F. Ali, D. Bell, and Mahir Arzoky, “Applications of Machine Learning in Education,” *Advances in educational technologies and instructional design book series*, pp. 72–113, January, 2021, doi: <https://doi.org/10.4018/978-1-7998-5009-0.ch004>.
69. ‌J. Stanley, Ronna ten Brink, A. Valiton, T. Bostic, and B. Scollan, “Chatbot Accessibility Guidance: A Review and Way Forward,” *Lecture notes in networks and systems*, pp. 919–942, September, 2021, doi: <https://doi.org/10.1007/978-981-16-1781-2_80>.
70. ‌Radka Nacheva and A. Jansone, “Heuristic Evaluation of AI-Powered Web Accessibility Assistants,” *Baltic Journal of Modern Computing*, vol. 11, no. 4, January, 2023, doi: <https://doi.org/10.22364/bjmc.2023.11.4.02>.
71. ‌A. P. Chaves and M. A. Gerosa, “How Should My Chatbot Interact? A Survey on Social Characteristics in Human–Chatbot Interaction Design,” *International Journal of Human–Computer Interaction*, vol. 37, no. 8, pp. 1–30, November, 2020, doi: <https://doi.org/10.1080/10447318.2020.1841438>.
72. S. Feuerriegel et al., "Generative AI," Business Information Systems Engineering, vol. 66, no. 1, 2024, pp. 111–126, various DOIs including <https://doi.org/10.1080/07421222.2023.2172773>.
73. Y. Cheng and H. Jiang, "How Do AI-driven Chatbots Impact User Experience? Examining Gratifications, Perceived Privacy Risk, Satisfaction, Loyalty, and Continued Use," Journal of Broadcasting & Electronic Media, Dec. 2020, https://doi.org/10.1080/08838151.2020.1834296&#8203
74. C.-C. Lin, A. Y. Q. Huang, and S. J. H. Yang, "A Review of AI-Driven Conversational Chatbots Implementation Methodologies and Challenges (1999–2022)," Sustainability, vol. 15, no. 4012, Feb. 2023, <https://doi.org/10.3390/su15054012&#8203>;
75. F. F.-H. Nah et al., "Generative AI and ChatGPT: Applications, Challenges, and AI-human Collaboration," Journal of Information Technology Case and Application Research, vol. 25, no. 3, 2023, https://doi.org/10.1080/15228053.2023.2233814&#8203;

# **12 Appendices**

APPENDIX A Source Code

APPENDIX B Gantts Chart

APPENDIX C Participant Information Sheet

APPENDIX D Information Consent Form

APPENDIX E Interview Questions

APPENDIX F Demographics Form