Bangla Voice-Based Conversational Chatbot Using NLP

Submitted By

M N Huda Nahid Khandaker
1935202075

Md. Monir Hossain
1935202084

A thesis report submitted in partial fulfillment of the requirements for the degree of Bachelor of Science in Computer Science and Engineering from City University.

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
CITY UNIVERSITY, DHAKA, BANGLADESH

JANUARY 2024

DECLARATION

This is to certify that the thesis titled “**Bangla Voice-Based Conversational Chatbot Using NLP”**– Using Machine Learning, Deep Learning, and Natural Language Processing (NLP) Based on Basic Banking Questionnaire” is the result of our study in partial fulfillment of the B.Sc. in Engineering degree under the supervision of Md Ataullah Bhuiyan, Senior Lecturer and Coordinator, Department of Computer Science and Engineering (CSE), City University, Dhaka, Bangladesh. It is also hereby declared that this project or any part of it has not been submitted elsewhere for the award of any degree.

Signature of Author’s

M N Huda Nahid Khandaker
ID: 1935202075
Department of Computer Science and
Engineering,
City University, Dhaka, Bangladesh

Md. Monir Hossain
ID: 1935202084
Department of Computer Science and
Engineering,
City University, Dhaka, Bangladesh

Signature of Supervisor

Md Ataullah Bhuiyan
Senior Lecturer and Coordinator
Department of Computer Science and
Engineering,
City University, Dhaka, Bangladesh

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to almighty Allah for his divine blessing in making it possible to complete this project successfully.

We would like to thank the following people for their help in the production of this project. We are deeply indebted to our supervisor **Md Ataullah Bhuiyan,** Senior Lecturer and Coordinator Department of Computer Science and Engineering, City University, Dhaka, Bangladesh. Without his help and support, throughout this project, it would not have been possible. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, and reading many inferior drafts and correcting them at all stages have made it possible to complete this project.

We would like to thank you from the bottom of our hearts to **Shaikh Shariful Habib**, Assistant Professor, and Head, Department of Computer Science and Engineering, Faculty of Science and Engineering, City University, for his precious advice, fruitful suggestions, and valuable assistance in completing my project. As well as our heartfelt gratitude to other faculty members and the staff of the Department of Computer Science and Engineering, City University, Dhaka, Bangladesh.

We are very grateful to all our faculty teachers who gave us their valuable guides to complete my graduation. We are also very grateful to all those people who have helped us to complete our project.

Finally, we must acknowledge with due respect the constant support and patience of our parents.

Abstract

Customers in the dynamic banking industry frequently have inquiries regarding several facets of their financial dealings. Given the significance of these interactions, more attention is being paid to the design of technologies that maximize the bank's time efficiency while simultaneously fostering a friendly and personable relationship with customers. The main goal of such a system is to build relationships that transcend beyond the transactional nature of banking. After that, a deep learning (DL) model and Bengali-specific machine learning (ML) techniques are incorporated into the Interactive Agent architecture to enable the system to comprehend and respond to user inquiries. Three modules make up the system: Text-to-Speech (TTS) Synthesis, Interactive Agent, and Automatic Speech Recognition (ASR). We tested the system's database using sentence transformers (paraphrase-mpnet-base-v2) and received satisfactory results. The TTS module uses the gTTS model. Furthermore, our findings indicated that an AI-based system could be a flexible option for any domain-specific reception system in charge of providing structured and effective customer service.

Keywords:- Interactive Agent, Automated speech recognition, Text-to-Speech Synthesis, Sentence Transformers.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

Chapter 1
Introduction

1.1 Introduction

The primary goals of Interactive Agents in the banking system are to improve client engagement and expedite routine inquiries and transactions, among other issues. Banks aim to increase customer satisfaction by offering prompt, individualized service through the use of interactive agents. Furthermore, interactive agents can help automate repetitive processes like account maintenance, fund transfers, and balance queries, freeing up human resources for more complicated problems. Additionally, this technology makes services available around-the-clock, guaranteeing that clients may get help whenever they need it [8].

By providing quick and easy access to fundamental banking services, Interactive Agent dramatically enhances the customer experience and increases satisfaction and loyalty [7]. In the end, it lowers expenses for the bank by improving operational efficiency and freeing up human agents to concentrate on more difficult and valuable duties. The advancement of speech recognition, computer vision, text to speech (TTS) synthesis, and natural language processing (NLP) provides the possibilities for the creation of automated reception management robots that are able to work as customer service receptionists [2].

Integrating Text-to-Speech (TTS) synthesis and Automatic Speech Recognition (ASR) in this study to simulate a human receptionist who is capable of working in Bengali language. The relevant questions at banks can be answered by the architecture utilized in a virtual receptionist robot. Three modules make up the system: Text-to-Speech (TTS) Synthesis, Interactive Agent, and Automatic Speech Recognition (ASR). Automatic Speech Recognition (ASR) created by Google researchers to train the model recognizes speech in Bengali. The encoding used for Bengali text is "paraphrase-mpnet-base-v2". gTTS is used in Text-to-Speech (TTS) responses to provide instantaneous human responses. These modules provide interactive, responsive agents in Bengali that have the potential to advance the current customer service model in banking reception.

1.2 Problem Statement

Customers in the dynamic banking industry frequently have inquiries regarding several facets of their financial dealings. Customers require full support for all account-related inquiries, transaction details, issues concerning various bank accounts, and worries regarding account management. Moreover, the range encompasses credit cards, loans, interest rates, and even more general questions about financial planning. Given the significance of these interactions, more attention is being paid to the design of technologies that maximize the bank's time efficiency while simultaneously fostering a friendly and personable relationship with customers. The main goal of such a system is to build relationships that transcend beyond the transactional nature of banking. The goal is to improve the customer experience by adding aspects of warmth and personalization to the customer contact process. This kind of approach not only answers questions quickly but also gradually fosters loyalty and trust. Moreover, expediting routine inquiries is the goal of deploying customer-focused technology, freeing up bank staff members' time and resources for other important tasks. Financial organizations can concentrate on areas like innovation, product development, and community involvement thanks to this efficiency gain, which eventually results in a more robust and extensive banking service. The proposed approach essentially aims to achieve a balance between operational efficiency and personalized customer involvement, making sure that both the bank and its clients gain from a mutually beneficial relationship.

1.3 Motivation

For numerous reasons, the thesis motivation for a Bengali voice-based chatbot in the financial sector is critical. For starters, it satisfies the requirement for inclusivity by catering to the linguistic diversity of Bengali speakers, making it more accessible to a wider user base. Second, it represents the banking industry's dedication to technological innovation, increasing efficiency and convenience through voice-based interactions. Furthermore, a Bengali voice-based chatbot is culturally appropriate, boosting user trust and engagement. Overall, the importance of adapting cutting-edge technology to local languages and cultural subtleties for a more inclusive and user-friendly banking experience in the Bangla-speaking community is emphasized by this thesis rationale.

1.4 Difficulties

There are various regions where Bengali is spoken, and each has its own dialect. Regional variations in the language's pronunciation, lexicon, and grammar make it challenging to create a voice-based system that can comprehend and react to a variety of dialects. Similar to numerous other languages, Bengali possesses nuanced tones, formalities, and politeness. It takes sophisticated natural language processing (NLP) skills and cultural sensitivity to build a voice-based chatbot that can comprehend and react to these subtleties. Bengali may have less voice data than other languages that are often spoken. A significant quantity of high-quality data is required for voice-based chatbot training, and a deficiency of this kind of data can make it difficult to build reliable and accurate models. Bengali, like other languages of South Asia, has its share of grammatical complexity. It has intricate sentence structures, a wide range of verb tenses, and a rich morphology. The development of a voice-based system that comprehends and responds grammatically requires a high level of language analysis expertise. Bengali speech recognition technology could not be as sophisticated as those of more widely spoken languages. Accurately transcribing spoken Bengali into text for further processing is challenging, particularly when handling accents and geographical variations. A culturally sensitive voice-based chatbot is essential. To deliver pertinent and appropriate answers, it is essential to comprehend regional idioms, customs, and cultural references. It could be challenging to imbue a chatbot with this cultural understanding. dealing with speech and text processing together is inherently more challenging than dealing with text alone. Advanced technology integration is required to ensure smooth speech and text format communication between the user and the system. Despite these challenges, advancements in speech recognition, machine learning, and natural language processing make it easier to create voice-based chatbots for languages like Bengali. Working together, linguists, technologists, and cultural specialists may overcome these challenges and create voice-based chatbot solutions that are both efficient and culturally appropriate.

1.5 Objectives

People are amid a technological revolution right now. They use technology everywhere in today's digital world. The ambitious goals of deploying a Bengali Chatbot system in the banking industry are multifaceted, aiming to improve important aspects of customer service, operational efficiency, and overall banking experiences. To begin, the system aims to change customer service by providing timely, precise, and tailored responses around the clock. This not only improves accessibility but also ensures that consumers receive individualized support, encouraging good and responsive engagement. Second, the emphasis is on operational efficiency, to streamline routine and repetitive processes through automation. As a result, the system contributes greatly to cost reduction and increases the overall efficiency of the bank's operations, allowing staff to focus on more delicate jobs that require human insight and strategic thinking.

- To develop a system that will be able to improve customer service by providing quick, accurate, and personalized responses with 24/7 accessibility.
- Improve operational efficiency to streamline routine and repetitive tasks. This contributes to cost reduction and increased efficiency in the bank's operation.
- Reduce operational costs and minimize the need for extensive human intervention in basic transactions.

1.6 Proposed Approach

In this thesis, several critical phases are included in the proposed approach to developing a Bengali voice-based chatbot for the banking sector. To begin, a thorough examination of the phonetics and linguistic intricacies of the Bengali language is required to achieve accurate and contextually relevant voice recognition. Following that, the chatbot's architecture should be created to include natural language processing (NLP) techniques suited to Bengali, allowing the system to successfully interpret and reply to user queries. Incorporating machine learning models can also improve the chatbot's ability to learn and adapt to user preferences over time. Collaborations with linguistic experts, as well as ongoing testing and refining, are essential for fine-tuning the chatbot's performance and ensuring it matches the specific needs and expectations of Bangla-speaking customers in the context of banking interactions.

1.7 Techniques

- Speech Recognition.
- Speech To Text.
- Text Similarity.
- Text To Speech.

1.8 Uniqueness

Increased internet penetration and smartphone usage have spurred the adoption of technology, including chatbots, within Bangladesh. Organizations across sectors like finance, e-commerce, and customer service are actively exploring chatbot integration to elevate user experiences and streamline communication.

1. **Bank-Centric Bangla Voice Chatbot:** The aim is to develop a Bengali voice-based conversational chatbot tailored for banks. It enables customers to efficiently conduct

banking tasks in Bengali, eliminating the need for physical bank visits or reliance on Internet banking.

2. **Operational Efficiency and Customer Satisfaction:** This initiative seeks to assist banks in curbing operational costs while enhancing customer satisfaction. By offering personalized and responsive services, it aims to improve overall customer experiences

1.9 Organization of the Study

The second chapter presented whatever has been accomplished in this arena. The second chapter's last portion demonstrates the depth that has resulted from their field's limitations. Finally, the research's major difficulties or hurdles are discussed.

The third chapter discusses the theoretical aspects of this study project. This chapter explains the quantitative methods used in this study to address the theoretical portion of the research. In addition, this chapter demonstrates the Deep Learning model and Machine Learning-based libraries.

The experimental findings, working procedure, and several outcomes are presented in Chapter Four. This chapter contains the deep learning (DL) model implementation for most possible outcomes in this project.

The fifth chapter presented the study's summarization, future development, and conclusion. This section is useful for enhancing the entire project report under the recommendations. The chapter concludes by demonstrating the limitations of this work.

Chapter 2

Literature Review

2.1 Related Works

Researchers applied various methods that incorporate Machine Learning (ML), Deep Learning (DL), and Natural Language Processing (NLP).

"Infini – A Keyword Recognition Chatbot"

Children with disabilities are equally entitled to a bright future with education being the cornerstone and a fundamental right for this achievement. Global organizations and policies are fighting for the inclusion of children with disabilities in mainstream schools. Nonetheless, inclusion is not enough if children's participation at school is not guaranteed. The provision of support services to ensure children's participation at school is an obligation of each state. Occupational therapists have a long tradition in the school context characterized by discrepancies. Therefore, this systematic literature review aims to map how school-based occupational therapists work with participation interventions to facilitate the participation of children with disabilities. After a comprehensive search in five databases, a selection process, quality assessment, and data extraction, this paper resulted in nine qualitative, quantitative, and mixed-design studies. The results presented under the scope of OTPF-3 and fPRC frameworks have revealed that occupational therapists traditionally use push-in and pull-out direct approaches as a medium to enhance the performance skills and activity competence of children. However, there is a shift of occupational therapists toward participation interventions in indirect ways through service delivery models focusing on the school's social and physical environment with their collaboration with teachers being of major importance [11].

"Human-robot-interaction using cloud-based speech recognition systems"

Emotion recognition, a key step of affective computing, is the process of decoding an embedded emotional message from human communication signals, e.g. visual, audio, and/or other physiological cues. It is well known that speech is the main channel for human communication and is thus vital in the signaling of emotion and semantic cues for the correct interpretation of contexts. In the verbal channel, the emotional content is largely conveyed as constant paralinguistic information signals, from which prosody is the most important component. The lack of evaluation of affect and emotional states in human-machine interaction is, however, currently limiting the potential behavior and user experience of technological devices. In this paper, speech prosody and related acoustic features of speech are used for the recognition of emotion from spoken Finnish. More specifically, methods for emotion recognition from speech relying on long-term global prosodic parameters are developed. An information fusion method is developed for short-segment emotion recognition using local prosodic features and vocal source features [2].

“Transformer based Bengali chatbot using general knowledge dataset”

Researcher's works on various transformer models, delivering big improvements in AI text models (NLP), are now being applied in Knowledge Tracing to track the knowledge of students over time. One of the first, SAINT, showed quite some improvement over the then SOTA results on the public EdNet dataset and caused an increase in research based on transformer-based models. In this paper, they aim to reproduce the SAINT results on the EdNet dataset but are unable to report a similar performance as the original paper. This might be due to implementation details, which we were not able to completely reconstruct. They pave the road for further reproducibility, as an increasingly important part of AI research. Furthermore, they apply the model to a company dataset much larger than any public dataset. Such a dataset is on the one hand more challenging (more skills mixed), and on the other hand, provides much more data (which should help our models). They compare the SAINT model and the seminal IRT model and find that the SAINT model performance is 4% better in AUC but 1.7% worse in RMSE. Our experiments on window size suggest that transformer models still struggle with modeling [3].

“An AI-based medical chatbot model for infectious disease prediction”

According to current evidence from animal studies, the phenotypes of the fetal inflammatory response are variable, ranging from spontaneous preterm birth to fetal death. The fetus is rather well protected against infectious agents by both structural and functional barriers. Toll-like receptors (TLR) are a part of the innate immune system. The binding of bacterial or viral components to TLR induces an inflammatory response in the host. The study addressed the hypothesis that the effects of bacterial lipopolysaccharide (LPS) on the fetus, depend on the route by which it reaches the fetus. Secondly, the role of the placenta in protecting the fetus from acute infectious challenges was evaluated. Additionally, the role of gestational age in the fetal immune response was studied. In the study, LPS from Gram-negative bacteria caused acute intrauterine inflammation in mice as indicated by the elevated levels of inflammatory mediators (e.g. cytokines) in amniotic fluid. The fetal heart revealed mRNA and protein expression of TLR4, which recognizes LPS. Moreover, the data showed a cytokine response in the fetal heart and severe cardiac dysfunction [4].

“Arabic educational neural network chatbot”

Super-resolution is a challenging problem of restoring details lost to diffraction in the image-capturing process. Degradations from the environment and the imaging device increase its difficulty, and they are strongly present in mobile phone cameras. The latest promising approaches involve convolutional neural networks, but little testing has been done on degraded images. Also, the sizes of neural networks raise a question of their applicability to mobile devices. A wide review of published super-resolution neural networks is done. Four of the network architectures are chosen, and their TensorFlow models are trained and tested for their output quality on high-quality and degraded images and compared against bicubic interpolation with sharpening. For the first time, MTF and CPIQ acute responses are measured from their outputs after processing photographs of a result [5].

“AI reception: An intelligent Bengali receptionist system integrating with face, speech, and interaction recognition”

This diploma research work aimed to create new applications for intelligent and adaptive lighting in retail environments. Lighting is an important factor in generating atmosphere in retail space and it has been shown to affect customer behavior. Combining intelligent technology with lighting design enables new applications for creating an environment that senses the presence of the user. It can be employed to adapt the lighting to inform and guide the customer by creating visual focal points. Alternatively, the level of illumination can be adapted to the different requirements of the use, e.g. the presence of customers or employees. The methods used in this study were scenario working and implementation. Four major themes were defined to approach the subject: 1. Navigation and guidance, 2. Product display and browsing, 3. Pleasure and entertainment, and 4. Natural light and simulated natural light [6].

“Chatbot design approaches for fashion E-commerce: an interdisciplinary review”

Software testing is an important part of software development projects. As the role of information technology (IT) becomes bigger and bigger in our everyday activities, it is clear that business operations and human well-being are dependent on information systems. To efficiently operate and run a business, companies reflect their processes to IT systems. A business process can cover many different organizational units, both in real life and in the IT system. Organizational units can have their separate IT system modules implemented, and data flows from module to module via interfaces. To ensure the correct functionality of the business process, end-to-end testing of the complete process across the IT systems is required. The test library is driven by a common keyword-driven test automation framework, Robot Framework [7].

“A model to develop chatbots for assisting the teaching and learning process”

The authors review existing literature regarding the teaching and learning of English pronunciation from the perspective of Finnish education. As English has become a worldwide language, in the future, people will need to be more adaptable to the versatile front of international spoken English, leading to the need for English as a foreign language, students to be able to understand and speak a comprehensible variation of it. Most students in Finland start learning English at the beginning of elementary school, and by the time they finish comprehensive school, they will be expected to understand various accents and speak an intelligible variation of English themselves. This, already, establishes the need for quality pronunciation education. In this thesis, schools’ explicit influence in the acquisition of pronunciation is looked at through the general viewpoint of foreign language learning in Finland, before considering the importance and intelligibility issues of pronunciation teaching and learning. This literature review aims to discover the methods and techniques used to teach English pronunciation to Finnish students, in addition to considering the various aspects affecting the optimization of learning [10].

“Shohojogi: An Automated Voice Chat System in the Bangla language for the banking system.”

Students come into the foreign language classroom with very different learning profiles, that is, readiness, interests, learning styles, social backgrounds, and emotional needs. To respond to these differences and needs, and consequently promote students' growth to their full potential as both human beings and FL users, teachers need to differentiate their teaching. This is also in the requirements of the Finnish Core Curriculum (2004) and the amendments of 2010, as differentiation is an integral means of providing general support for all students. Research has been done on differentiating foreign languages, but there is space for a more encompassing study exploring differentiation in the FL classroom in the comprehensive schooling system in general. This study aims at increasing understanding and knowledge on differentiation in the foreign language classroom as a phenomenon and spreading awareness of diverse, fruitful, and respectful means of differentiation that FL teachers could apply in their everyday work. In other words, the main question of their study is what should and could differentiation in the FL classroom be like [8].

"Human-Robot Interaction in Bengali language for Healthcare Automation integrated with Speaker Recognition and Artificial Conversational Entity"

They describe the current state of artificial intelligence, and how it can be applied to the industry. Artificial intelligence is considered one of the major enablers of a movement towards the fourth industrial revolution. It has the potential to increase growth across our global economy, for instance, by automating work and enhancing the performance of humans. It can also be implemented to optimize decision-making and create interconnected supply chains, where the availability of real-time information from multiple facets enables complete optimization and transparency, which would not be attainable using only currently established methods. While we are yet far from a point where AI could completely replace human workers on a larger scale, it can still be useful for automating certain work activities, which can be identified through an appropriate categorization. In addition to determining current technologies' potential in the automation of work, it explains how AI can be used to create a loop between the physical and digital worlds. In practice, this means creating optimized systems, in which continuous information flow and algorithmic decision-making enable completely new levels of operational efficiency [9].

2.2 Comparison Table with the Existing Work

Features	Survey on Intelligent Chatbots: State-of-the-Art and Future Research Directions	Human-robot-interaction using cloud-based speech recognition systems	Transformer-based Bengali chatbot using general knowledge dataset	AnAI-based medical chatbot model for infectious disease prediction	Arabic Educational Neural Network Chatbot	Bangla Voice-Based Conversational Chatbot Using NLP (This Study)
Detect Human Speech	No	Yes	No	No	No	Yes
Convert speech into text	No	No	No	No	No	Yes
Understand the questions	No	Yes	Yes	Yes	Yes	Yes
Convert answers text to speech	No	No	No	No	No	Yes
Deliver an answer in a speech	No	Yes	No	No	No	Yes

Table 2.1: Comparing with existing work

Chapter 3

Methodology

3.1 Overview

This section demonstrates the robust architecture of “Bangla Voice-Based Conversational Chatbot Using NLP”. All the modules developed with multiple deep learning (DL) and natural language processing (NLP) techniques.

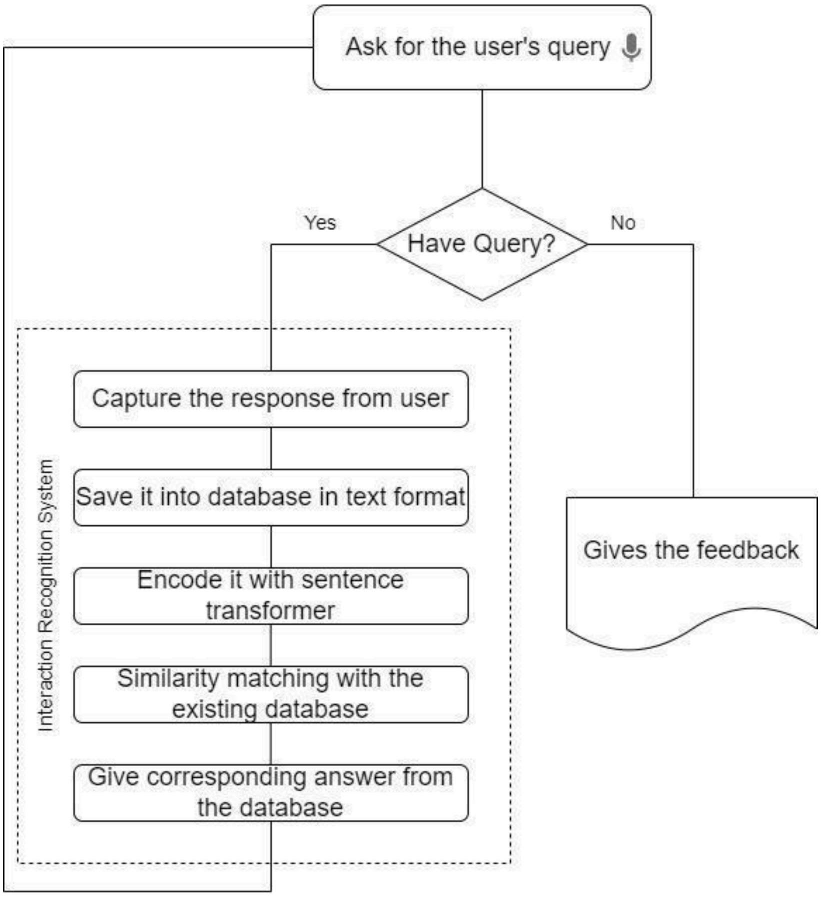

Ask for the user's query
Have Query?
Yes
No
Interaction Recognition System
Capture the response from user
Save it into database in text format
Encode it with sentence transformer
Similarity matching with the existing database
Give corresponding answer from the database
Gives the feedback

Figure 3.1.1: System architecture of the interactive agent

In the beginning, the user asks the query. If the query matches the database, it will go to the module otherwise the system gives feedback to the user. After the system catches the response from the user which is shown in figure 3.1.1, it goes to the database in text format by Text-to-Speech (TTS) synthesis. The text data is encoded by “paraphrase-mpnet-base-v2” sentence transformer. The system matches the similarity of encoded data by cosine similarity with the existing database. After similarity matching the system gives the most corresponding answer from the database. All these processes are made possible by the Interaction Recognition System.

3.2 Interaction Recognition System

Developing the interactive agent is enriched with ASR (Automatic Speech Recognition), Interactive Agent, and TTS (Text-to-Speech) synthesis module shown in figure 3.2.1. The system captures the user's voice through a microphone and Automatic Speech Recognition (ASR) converts the speech into the Bangla text format using gTTS. Afterward, the converted text goes to the database for encoding the text data with "paraphrase-mpnet-base-v2" sentence transformer model. In the database using cosine similarity to match the embedded data from the predefined question and provided the corresponding answers from the database. Finally, the TTS system transforms the text response of the interactive agent into speech.

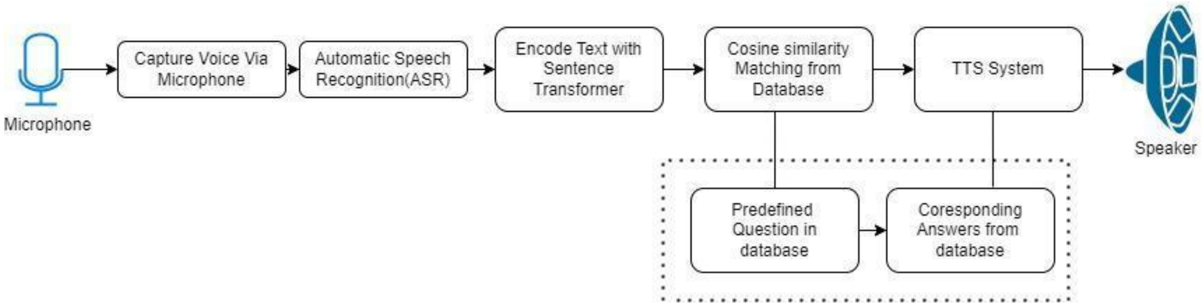

Microphone
Capture Voice Via Microphone
Automatic Speech Recognition(ASR)
Encode Text with Sentence Transformer
Cosine similarity Matching from Database
TTS System
Speaker
Predefined Question in database
Coresponding Answers from database

Figure 3.2.1: Workflow of the proposed method

3.3 Interactive Agent in Bengali

The main task of the Interaction Recognition system is to receive a question as a text from the ASR system, then give the most relevant answer from the knowledge-based database. This Interactive Agent is mainly designed for the banking customer care services. The database is created with bank related questions with corresponding answers in Bengali language. A sentence transformers model "paraphrase-mpnet-base-v2" has been used to perform the sentence embedding the sentence. The system used sentence transformer; a python-based framework that offers miscellaneous tunable pre-trained models trained on the various datasets. As seen in Figure. 3.2.1, the customized Interactive Agent can be classified into five steps.

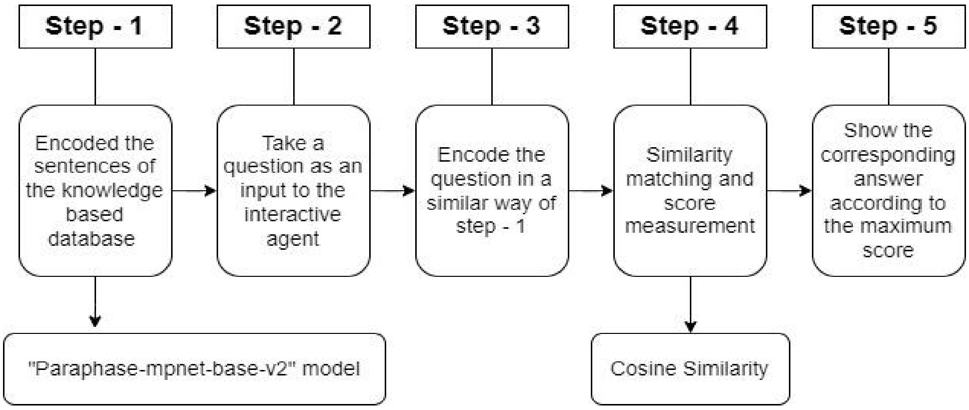

Step - 1
Step - 2
Step - 3
Step - 4
Step - 5
Encoded the sentences of the knowledge based database
Take a question as an input to the interactive agent
Encode the question in a similar way of step - 1
Similarity matching and score measurement
Show the corresponding answer according to the maximum score
"Paraphase-mpnet-base-v2" model
Cosine Similarity

Figure 3.3.1: Steps of Interactive Agent

Step 1: Encode all of the database's sentences: Sentence Transformer, a Python-based framework, is used to encode all of the sentences in the knowledge-based database. The database is divided into two sections: the question portion and the answer section. For sentence embedding, the "paraphrase-mpnet-base-v2" model was employed, which turns each standardized question and associated answer into a vector format.

Step 2: Consider the following query as input to the Interactive Agent: The process that follows takes a question from a system user. ASR accepts voice input from a user with bank-related questions. The text is then converted into Bangla using the "paraphrase-mpnet-base-v2" model and sent to the Interactive Agent.

Step 3: Encode the following question: It is necessary to encode the sentence received from users. For sentence embedding, the same approaches as in step 1 were used.

Step 4: Similarity Matching & Score Measurement: Counting frequent terms between the database's question section and user input could be a solution for constructing a closed-domain Interactive Agent. However, problems will arise if the size of the prepared texts is too high because the number of common terms increases with database size. The cosine similarity metric is used to get the best score between the user's question and questions from the database.

Step 5: Show the corresponding answer based on the highest possible score: The system displays the corresponding response to that question by computing the maximum score using cosine similarity between the prepared questions and user input. The Interactive Agent's output stays in text format, which serves as the input for text-to-speech synthesis.

3.4 Text-to-Speech Synthesis

To develop a voice-based conversational agent that can respond to different questions about banking queries, the response's output needs to be audible. To translate text input into speech at first, by employing the "gTTS" module, which is based in Python. Python 2 and 3 versions both support this library. The primary justification for utilizing this library in the developed system is that, out of all the TTS-based Python libraries available, "gTTS" provides the highest quality sound. Moreover, the text response's speech production speed is very remarkable.

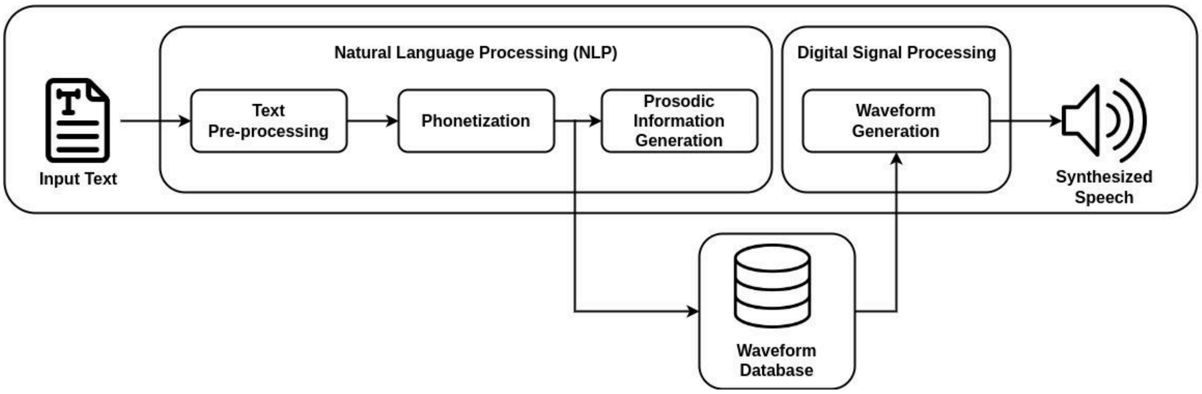


Figure 3.4.1: Text-to-Speech (TTS) synthesis

3.5 Experimental Setup

Developing the “Bangla Voice-Based Conversational Chatbot Using NLP " system needs some experimental setup.

3.5.1 Artificial Intelligence (AI)

The science of creating machines with human-like thought processes is known as artificial intelligence. It is capable of actions deemed "smart." Through the use of an AI-driven algorithm that effectively searches customer care records and previous exchanges for language patterns resembling the first query, chatbot programs aim to delight customers. This enables it to swiftly and precisely provide the most relevant response.

3.5.2 Machine Learning (ML)

Within the fields of computer science and artificial intelligence (AI), machine learning focuses on using data and algorithms to simulate human learning processes and progressively increase their accuracy. The most advanced chatbots are machine-learning chatbots or artificial intelligence (AI) chatbots. They let consumers pose complex, free-form queries and provide the most organic answers. Over time, these chatbots get better at responding as they keep picking up new skills from discussions.

3.5.3 Deep Learning (DL)

Some of the pre-processing of data that is usually required for machine learning is eliminated by deep learning. These algorithms automate feature extraction, reducing the need for human specialists, and can ingest and handle unstructured data, such as text and photos. Computers and chatbots can better understand these interrelated meanings with the aid of deep learning. Deep learning belongs to a class of machine learning techniques that imitate the functioning of a human brain network. It mimics the information-sharing between brain neurons in a semantic network.

3.5.4 Natural Language Processing (NLP)

A machine-learning technique called natural language processing (NLP) enables computers to understand, manipulate, and interpret human language. A chatbot that uses natural language processing is a computer program that can comprehend and react to spoken language. NLP-powered bots enable users to interact with computers in a way that feels authentic and human, emulating face-to-face interactions.

3.5.5 Automatic Speech Recognition (ASR)

Automatic Speech Recognition (ASR) is a technology that converts spoken language into written text. It is also commonly referred to as speech-to-text or voice recognition. ASR systems have a wide range of applications, including transcription services, voice-activated virtual assistants, voice command systems, and more.

3.5.6 Sentence Transformers

Sentence transformers are a specific type of deep learning model designed to reliably preserve the semantic meaning of text by translating sentences or other text into fixed-length numerical representations (embeddings). Texts are embedded in a vector space so that related texts are near to each other. This allows for the use of applications like retrieval, grouping, and semantic search.

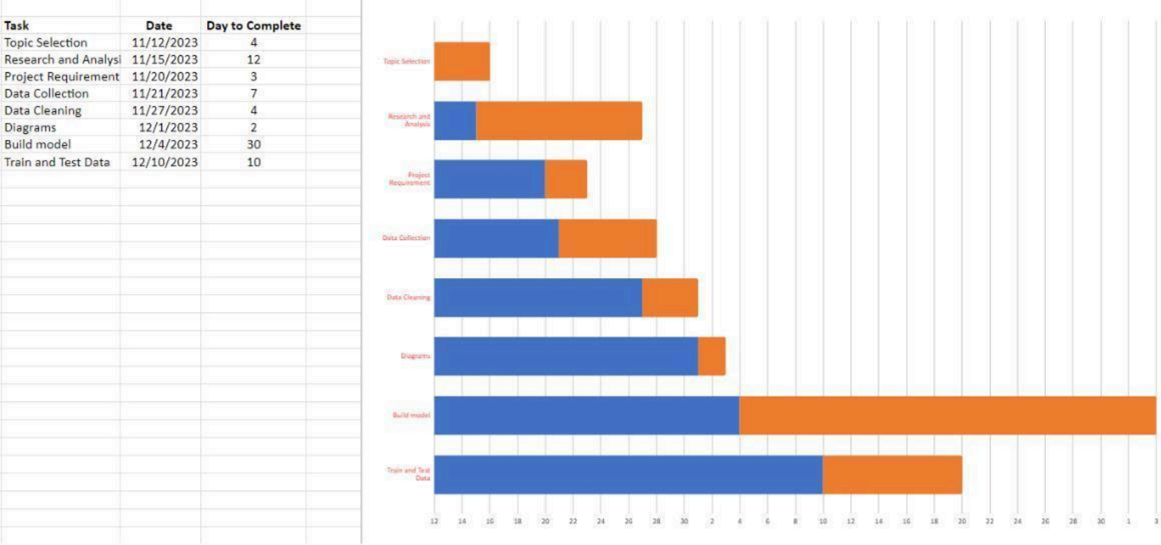
3.5.7 Cosine Similarity

Cosine similarity is a metric that is used in a variety of machine learning algorithms, such as the KNN for calculating the distance between neighbors, recommendation systems for matching movies, and textual data for comparing texts within a document. It offers a reliable method for comprehending the semantic similarity of datasets, papers, or pictures. For instance, vector search engines frequently use cosine similarity to identify the records that are most pertinent to a particular query, streamlining and improving search operations.

3.6 Dataset Description

The thesis criteria are not met by any Bangla datasets on Kaggle or other online resources. So, the dataset was collected from First Security Islami Bank. We visited the bank multiple times (21 November to 23 November). We met the manager and two other employees and a number of customers to collect the query and answers. The dataset is based on multiple FAQ questions and answers which is a conversion that takes place between customer and banker. We collect about 152 queries and answers for our dataset. The dataset is in the CSV format that is stored in the database. We tested the dataset using sentence transformers "paraphrase-mpnet-base-v2" model which turns each standardized question and associated answer into a vector format. We employed a machine learning algorithm called cosine similarity to find the similarity from the dataset based on the given query from the user.

3.7 Gantt Chart



Task	Date	Day to Complete
Topic Selection	11/12/2023	4
Research and Analysi	11/15/2023	12
Project Requirement	11/20/2023	3
Data Collection	11/21/2023	7
Data Cleaning	11/27/2023	4
Diagrams	12/1/2023	2
Build model	12/4/2023	30
Train and Test Data	12/10/2023	10

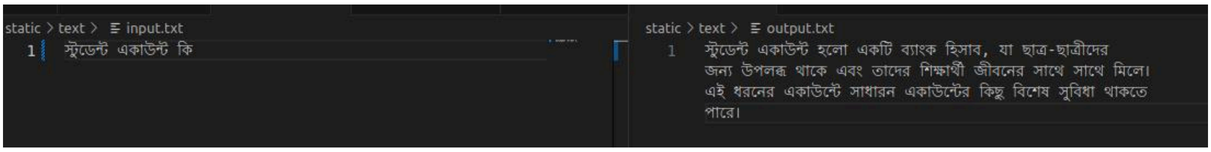
Figure 3.6.1: Gantt Chart

Chapter 4

Result and Discussion

4.1 Result Analysis

This interactive agent transforms human vocal to text data and text data to human vocal using gTTS. The system encodes the query using sentence transformers and finds the best possible answer from the database using cosine similarity. The figure 4.1.1 shows the possible outcome of the right query.

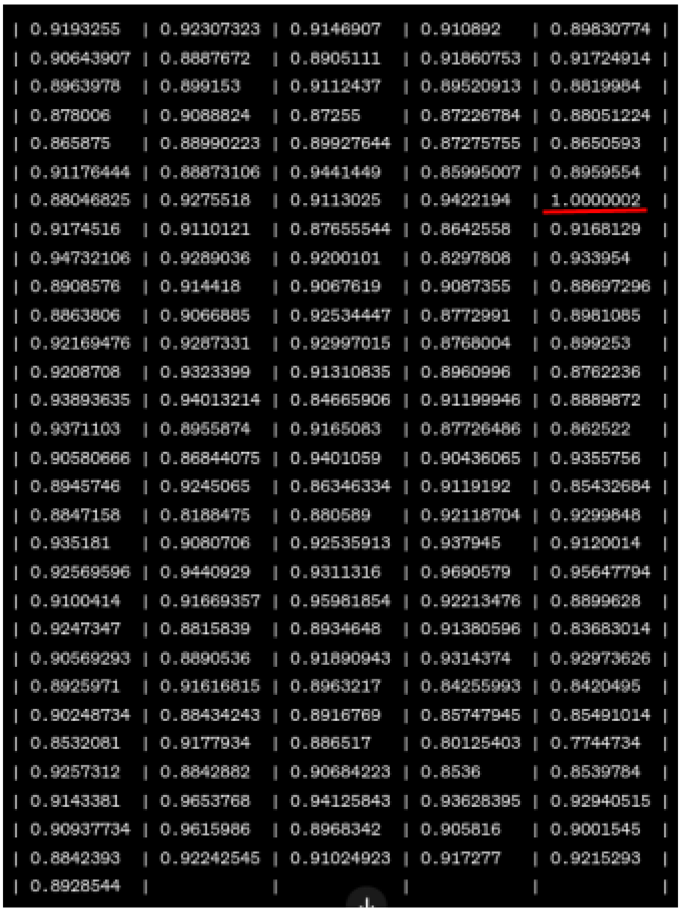


স্টুডেন্ট একাউন্ট কি

স্টুডেন্ট একাউন্ট হলো একটি ব্যাংক হিসাব, যা ছাত্র-ছাত্রীদের জন্য উপলব্ধ থাকে এবং তাদের শিক্ষার্থী জীবনের সাথে সাথে মিলে। এই ধরনের একাউন্টে সাধারন একাউন্টের কিছু বিশেষ সুবিধা থাকতে পারে।

Figure 4.1.1 Input and output of right query

The figure shows the most possible outcome from the interactive agent. The system tested with several queries and it gives the similarity score of each queries. In figure 4.1.2 shows the similarity score of the given query in figure 4.1.1.

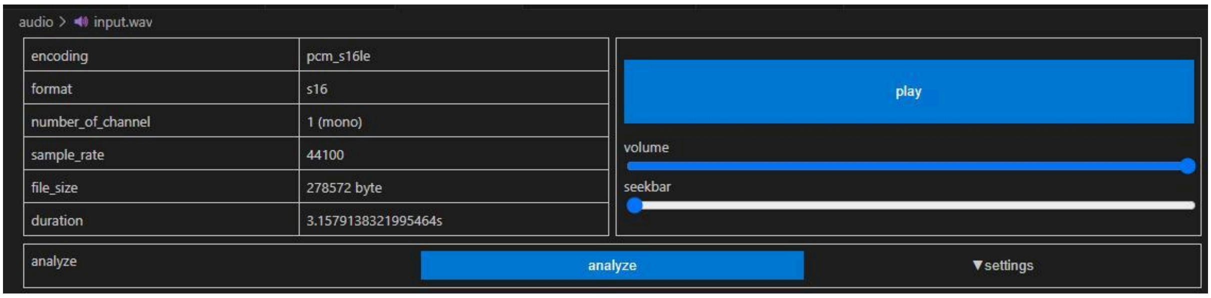


```
| 0.9193255  | 0.92307323 | 0.9146907  | 0.910892   | 0.89830774 |
| 0.90643907 | 0.8887672  | 0.8905111  | 0.91860753 | 0.91724914 |
| 0.8963978  | 0.899153   | 0.9112437  | 0.89520913 | 0.8819984  |
| 0.878006   | 0.9088824  | 0.87255    | 0.87226784 | 0.88051224 |
| 0.865875   | 0.88990223 | 0.89927644 | 0.87275755 | 0.8650593  |
| 0.91176444 | 0.88873106 | 0.9441449  | 0.85995007 | 0.8959554  |
| 0.88046825 | 0.9275518  | 0.9113025  | 0.9422194  | 1.0000002  |
| 0.9174516  | 0.9110121  | 0.87655544 | 0.8642558  | 0.9168129  |
| 0.94732106 | 0.9289036  | 0.9200101  | 0.8297808  | 0.933954   |
| 0.8908576  | 0.914418   | 0.9067619  | 0.9087355  | 0.88697296 |
| 0.8863806  | 0.9066885  | 0.92534447 | 0.8772991  | 0.8981085  |
| 0.92169476 | 0.9287331  | 0.92997015 | 0.8768004  | 0.899253   |
| 0.9208708  | 0.9323399  | 0.91310835 | 0.8960996  | 0.8762236  |
| 0.93893635 | 0.94013214 | 0.84665906 | 0.91199946 | 0.8889872  |
| 0.9371103  | 0.8955874  | 0.9165083  | 0.87726486 | 0.862522   |
| 0.90580666 | 0.86844075 | 0.9401059  | 0.90436065 | 0.9355756  |
| 0.8945746  | 0.9245065  | 0.86346334 | 0.9119192  | 0.85432684 |
| 0.8847158  | 0.8188475  | 0.880589   | 0.92118704 | 0.9299848  |
| 0.935181   | 0.9080706  | 0.92535913 | 0.937945   | 0.9120014  |
| 0.92569596 | 0.9440929  | 0.9311316  | 0.9690579  | 0.95647794 |
| 0.9100414  | 0.91669357 | 0.95981854 | 0.92213476 | 0.8899628  |
| 0.9247347  | 0.8815839  | 0.8934648  | 0.91380596 | 0.83683014 |
| 0.90569293 | 0.8890536  | 0.91890943 | 0.9314374  | 0.92973626 |
| 0.8925971  | 0.91616815 | 0.8963217  | 0.84255993 | 0.8420495  |
| 0.90248734 | 0.88434243 | 0.8916769  | 0.85747945 | 0.85491014 |
| 0.8532081  | 0.9177934  | 0.886517   | 0.80125403 | 0.7744734  |
| 0.9257312  | 0.8842882  | 0.90684223 | 0.8536     | 0.8539784  |
| 0.9143381  | 0.9653768  | 0.94125843 | 0.93628395 | 0.92940515 |
| 0.90937734 | 0.9615986  | 0.8968342  | 0.905816   | 0.9001545  |
| 0.8842393  | 0.92242545 | 0.91024923 | 0.917277   | 0.9215293  |
| 0.8928544  |            |            |            |            |
```

Figure 4.1.2: Similarity Score

In the figure 4.1.2 the red mark notifies the highest similarity score from the dataset after checking the similarity according to the given query.

Also the given query saves in input.txt file as text and the speech saves as audio file. After every query those files are replaced with new text and audio. In figure 4.1.1 also shows the saved text file.

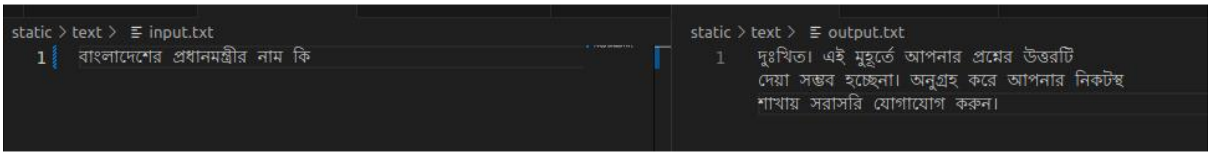


audio > input.wav

encoding	pcm_s16le
format	s16
number_of_channel	1 (mono)
sample_rate	44100
file_size	278572 byte
duration	3.1579138321995464s

Figure 4.1.3: Saves the audio

Besides, the user also asks many wrong queries which are not relevant with the predefined database. So, the system gives feedback to the user.



static > text > input.txt

```
বাংলাদেশের প্রধানমন্ত্রীর নাম কি
```

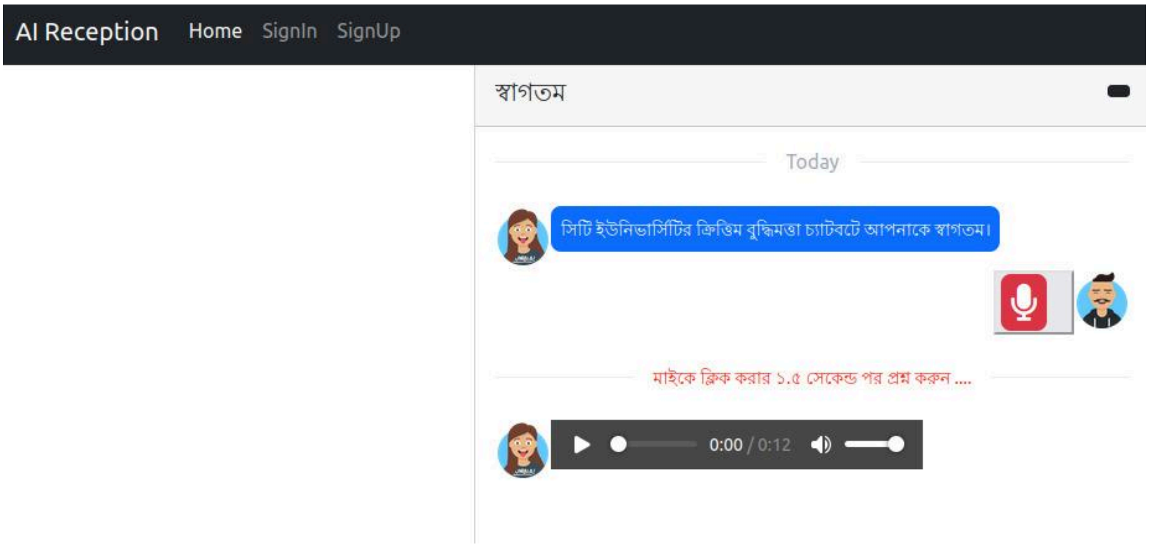
static > text > output.txt

```
দুঃখিত। এই মুহূর্তে আপনার প্রশ্নের উত্তরটি
দেয়া সম্ভব হচ্ছেনা। অনুগ্রহ করে আপনার নিকটস্থ
শাখায় সরাসরি যোগাযোগ করুন।
```

Figure 4.1.4: Feedback of the wrong query

4.2 Web Application of the interactive agent

The interactive agent has a web application. To make this system dynamic, the developers used flask to develop a web application for the interactive agent. Flask is a small and lightweight Python web framework that provides useful tools and features that make creating web applications in Python easier. The figure 4.1.5 shows the user interface of the web application. In the navbar it's called an AI Reception. After opening the application, the user can see a welcome message and a microphone button. When asking queries, the user should press the microphone button and wait for 1.5 seconds. After waiting for a while the user can query whatever they want and the interactive agent responds with the most possible answer in speech. The system saves the answer in speech that will be listened to before another query asked by the user.



সিটি ইউনিভার্সিটির কৃত্রিম বুদ্ধিমত্তা চ্যাটবটে আপনাকে স্বাগতম।

মাইকে ক্লিক করার ১.৫ সেকেন্ড পর প্রশ্ন করুন ....

Figure 4.2.1: Web application

4.3 Discussion

This interactive agent transforms human vocal to text data and text data to human vocal using gTTS. The system encodes the query using sentence transformers and finds the best possible answer from the database using cosine similarity. The system tested with several queries and it gives the similarity score of each queries. It saves the given query as a text file and the speech saves as an audio file. After finding the answer according to the given query from the dataset, the system gives the response as speech. Sometimes the system doesn't recognize the vocal which is given to the system then the system gives a feedback to the user as a speech form. Also given feedback when the system when some inappropriate queries are asked by the user. Afterwards developing a web application for the interactive agent to make the system dynamic.

Chapter 5
Conclusion and Future Work

5.1 Conclusion

An AI based interactive agent for the banking reception system that transforms to ensure user satisfaction by providing voice-enabled question answering in Bengali language. In addition, the system comprises three essential modules: Automatic Speech Recognition (ASR), an Interactive Agent, and Text-to-Speech (TTS) Synthesis. The ASR module utilizes a model developed by Google researchers to accurately transcribe Bengali speech. The Interactive Agent is empowered by the "paraphrase-mpnet-base-v2" to encode Bengali text, enhancing its ability to comprehend and respond to user queries. Finally, the Text-to-Speech module utilizes gTTS to generate human-like responses instantly. Together, these integrated modules form a responsive interactive agent, representing a significant advancement in modern-day customer care systems for banking reception, with the capability to efficiently handle and address customer queries in the Bengali language.

5.2 Future Work

In future, enhancing the virtual receptionist system for Bengali language customer service in banks could involve refining the Automatic Speech Recognition (ASR) module through continuous training with a diverse dataset to improve accuracy and adaptability to various accents and speech patterns. Additionally, incorporating advanced natural language processing (NLP) techniques specific to Bengali, along with sentiment analysis, could enable the Interactive Agent module to comprehend and respond more effectively to nuanced customer queries. Further exploration into deep learning architectures for Text-to-Speech (TTS) synthesis tailored to Bengali, leveraging expressive and context-aware voice generation, would contribute to creating more natural and human-like responses. Integration with evolving technologies, such as emotional intelligence algorithms, could also enhance the system's ability to perceive and respond empathetically to customer emotions, ultimately refining the overall customer care experience in the banking sector.

References

[1] E. H. Almansor and F. K. Hussain, “Survey on intelligent chatbots: State-of-the-art and future research directions,” in *Advances in Intelligent Systems and Computing*, Cham: Springer International Publishing, 2020, pp. 534–543.

[2] C. Deuerlein, M. Langer, J. Seßner, P. Heß, and J. Franke, “Human-robot-interaction using cloud-based speech recognition systems,” *Procedia CIRP*, vol. 97, pp. 130–135, 2021.

[3] A. K. M. Masum, S. Abujar, S. Akter, N. J. Ria, and S. A. Hossain, “Transformer based Bengali chatbot using general knowledge dataset,” in *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 2021.

[4] S. Chakraborty *et al.*, “An AI-based medical chatbot model for infectious disease prediction,” *IEEE Access*, vol. 10, pp. 128469–128483, 2022.

[5] B. A. Alazzam, M. Alkhatib, K. Shaalan, and A. Alazzam, “Arabic educational neural network chatbot,” *Inf. Sci. Lett.*, vol. 12, no. 6, pp. 2579–2589, 2023.

[6] R. A. Nabid, S. I. Pranto, N. Mohammed, F. Sarker, M. N. Huda, and K. A. Mamun, “AI reception: An intelligent Bengali receptionist system integrating with face, speech, and interaction recognition,” in *Bangabandhu and Digital Bangladesh*, Cham: Springer International Publishing, 2022, pp. 76–91.

[7] A. R. D. B. Landim *et al.*, “Chatbot design approaches for fashion E-commerce: an interdisciplinary review,” *Int. J. Fash. Des. Technol. Educ.*, vol. 15, no. 2, pp. 200–210, 2022.

[8] K. A. R. Arnab, I. Nabi, and M. I. Hossain, *Shohojogi: An Automated Voice Chat System in the Bangla language for the banking system*. .

[9] S. I. Pranto *et al.*, “Human-Robot Interaction in Bengali language for Healthcare Automation integrated with Speaker Recognition and Artificial Conversational

Entity," in *2021 3rd International Conference on Electrical & Electronic Engineering (ICEEE)*, 2021.

[10] S. Mendoza, L. M. Sánchez-Adame, J. F. Urquiza-Yllescas, B. A. González-Beltrán, and D. Decouchant, "A model to develop chatbots for assisting the teaching and learning process," *Sensors (Basel)*, vol. 22, no. 15, p. 5532, 2022.

[11] M. D. Thakkar, C. U. Sanghavi, M. N. Shah, and N. Jain, "Infini – A Keyword Recognition Chatbot," in 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), 2021.

[12] S. Quarteroni and S. Manandhar, "A chatbot-based interactive question answering system," Decalog, vol. 83, 2007.

[13] R. Burri, "Improving user trust towards conversational chatbot interfaces with voice output," 2018.

[14] P. Klaus and J. Zaichkowsky, "AI voice bots: a services marketing research agenda," J. Serv. Mark., vol. 34, no. 3, pp. 389–398, 2020.