

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/383849671>

An Artificial Intelligence Powered Bengali Voice Based Conversational Chatbot for Banking Sectors

Preprint · September 2024

DOI: 10.13140/RG.2.2.17082.20169

CITATIONS

0

READS

116

5 authors, including:



Md. Taufiqul Haque Khan Tusar

LaLoka Labs

11 PUBLICATIONS 67 CITATIONS

SEE PROFILE



Md Foysal

City University

1 PUBLICATION 0 CITATIONS






SEE PROFILE

An Artificial Intelligence Powered Bengali Voice Based Conversational Chatbot for Banking Sectors

¹M N Huda Nahid Khandaker*, ²Md. Monir Hossain, ³Md. Ataullah Bhuiyan, ⁴Md. Taufiqul Haque Khan Tusar, ⁵Md. Foysal

Dept. of Computer Science and Engineering, City University, Dhaka 1340, Bangladesh

¹nahid.cse52@gmail.com, ²monir52cse.cu@gmail.com, ³abab_bn@yahoo.com, ⁴taufiqkhanatusar@gmail.com, ⁵mdfoysal.cysec@gmail.com

¹0009-0004-8965-5712 , ²0009-0000-1114-2109 , ³0009-0001-7975-7553 , ⁴0000-0002-5586-6819 , ⁵0009-0005-3614-3569 

Abstract—Customers in the dynamic banking industry frequently have inquiries regarding several facts of their financial dealings. The main goal of our system is to build relationships that transcend beyond the transactional nature of banking. To enable the system to understand and reply to user requests, we employ a deep learning (DL) model and integrate machine learning (ML) techniques particular to Bengali into the Interactive Agent architecture. Three modules make up the system: Text-to-Speech (TTS) Synthesis, Interactive Agent, and Automatic Speech Recognition (ASR). We trained the interactive agent using sentence transformers (paraphrase-mpnet-base-v2) model to find the accurate answer from the database. We use the gTTS model for Text-to-Speech (TTS) synthesis. 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.

I. INTRODUCTION

Interactive Agents in the banking system are to improve client engagement and expedite routine inquiries and transactions, among other issues [13]. 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 [2]. 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, TTS synthesis, and natural language processing (NLP) provides the possibilities for the creation of automated reception management systems that are able to work as customer service receptionists. Integrating TTS synthesis and ASR in this study to simulate a human receptionist who is capable of working in Bengali language [6] [9]. Three modules

make up the system: TTS Synthesis, Interactive Agent, and ASR. ASR created by Google researchers to train the model recognizes speech in Bengali. The interactive agent trained with "paraphrase-mpnet-base-v2" model. The gTTS is used in TTS synthesis 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.

Customers in the dynamic banking industry frequently have inquiries regarding several facets of their financial dealings [10]. In the midst of a contemporary technological revolution, individuals actively engage with technology in diverse contexts within the digital milieu. The ambitious objectives of implementing a Bengali Chatbot system in the banking sector are multifaceted, targeting enhancements in customer service, operational efficiency, and overall banking experiences. Firstly, the system seeks to revolutionize customer service by furnishing timely, precise, and tailored responses consistently, enhancing accessibility and encouraging responsive engagement [11]. Secondly, there is a focus on optimizing operational efficiency through automation, leading to substantial cost reduction and heightened operational efficacy, enabling staff to concentrate on tasks requiring human insight and strategic acumen [12]. The linguistic landscape of Bengali is characterized by diverse regional dialects, presenting a formidable challenge for the development of a voice-based chatbot. Variations in pronunciation, lexicon, and grammar necessitate sophisticated NLP capabilities and cultural acumen. Insufficient high-quality voice data further complicates model training, while the language's grammatical intricacies demand advanced language analysis expertise. Despite these challenges, advancements in speech recognition, machine learning, and NLP offer opportunities for collaborative efforts among linguists, technologists, and cultural specialists to create culturally sensitive and efficient voice-based chat-bot solutions.

II. LITERATURE REVIEW

Researchers applied various methods that incorporate ML, DL, and NLP.

In the study conducted by [1], Human-Computer Interaction (HCI) is a pivotal field exploring the dynamic between humans and machines. Dialog systems, such as chatbots, voice interfaces, and personal assistants, exemplify HCI applications designed for natural language interaction. Chatbots, facilitating information retrieval, are increasingly integral in automating customer service. As the demand for automated services rises, the complexity of developing intelligent bots capable of human-level responses presents a substantial challenge. This paper conducts a state-of-the-art survey on chatbot methodologies, focusing on response generation proficiency, identifying pertinent research issues and challenges, thereby guiding future endeavors.

In [2], advancements in NLP have significantly enhanced the capabilities of speech processing systems, primarily leveraging machine learning technologies. This study contributes by developing a software interface facilitating voice control for a lightweight robot through cloud-based speech recognition. The focus lies in recognizing commands, converting them into machine-readable code, and evaluating diverse cloud services for robot control. The implemented control architecture incorporates fundamental features of cloud-based speech processing, exemplified in a proof-of-concept application enabling users to command robot movements through speech, even in the presence of background noise, ensuring commendable process reliability.

The authors [3], leveraging deep neural models, demonstrates remarkable proficiency after training on a comprehensive dataset for an AI chatbot. This study emphasizes the superiority of deep neural models, particularly highlighting the effectiveness of the seq2seq learning paradigm, implemented through recurrent neural networks (RNNs). While RNNs traditionally address sequence-related tasks, the introduction of attention mechanisms within the encoder and decoder significantly enhances seq2seq model performance. Transitioning to transformer models, these advanced architectures, featuring multiple attention mechanisms, outperform RNN-based models, notably reducing training time and achieving state-of-the-art results in sequence transduction. In the study, they applied the transformer model to develop a Bengali general knowledge chatbot, attaining an 85.0 BLEU score on the Bengali general knowledge Question Answer (QA) dataset. To validate the transformer model's superiority, they trained a seq2seq model with attention on our dataset, yielding a significantly lower BLEU score of 23.5.

The authors [4], aims to succinctly outline strategies for advancing the integration of chatbots in the medical domain to address infectious diseases. Their approach involves raising user awareness and delivering tailored medical solutions for disease prevention. Utilizing natural language processing, they analyze human behaviors, shaping an AI Chatbot interaction and prediction model through a deep feedforward multilayer perceptron. Their study identifies a knowledge gap in theoretical guidelines for creating lifestyle improvement AI chatbots. The proposed model demonstrates a minimum loss of 0.1232 and achieves a peak accuracy of 94.32%, showcasing its

potential applications in medical chatbots amid health crises, particularly pandemics like COVID-19.

The study [5] proposed Artificial Intelligence (AI)-driven chatbots, or machine-driven conversational systems, have become increasingly popular, demonstrating their capacity to mimic human communication in a variety of natural languages. But while chatbots have been reported to succeed in widely spoken languages, the predicted success of chatbots in Arabic has not materialized. There are now scholarly initiatives aimed at tackling this issue, especially with regard to improving the use of Arabic in academic contexts. The goal of their project is to create an instructional chatbot in Arabic using datasets from Arabic-language educational websites and NLP techniques. They collected pertinent literature, trained the system using a neural network model for the chatbot. Therefore, when it comes to answering questions about educational laws in the United Arab Emirates, their Arabic chatbot is quite skilled.

In [6], the AI-based Smart Reception, uses OpenFace for facial identification and achieves 92.92% accuracy in a surprisingly short 1×10^{-5} second training period. Text-to-Speech (TTS) synthesis, Interactive Agent, and Automatic Speech Recognition (ASR) modules are all integrated into the Interaction Recognition system. The Word Error Rate (WER) of 42.15% is achieved by the Deep Speech 2 model using OpenSLR-Large Bengali ASR Training Data. With a noteworthy 92% accuracy, the three-step evaluation of the Interaction Recognition system using the BERT sentence transformer produced satisfactory results. The versatility and effectiveness of AI-based systems in augmenting domain-specific reception systems for thorough and effective offline and online customer care are highlighted by this research.

In study [7], fashion firms use chatbots for individualized consumer interactions in the rapidly expanding e-commerce space. Chatbots provide a revolutionary edge in online assistance and client engagement. Existing research examines aspects of consumer behavior and technology, but this work offers a new interdisciplinary viewpoint by thoroughly classifying recent studies to inform future research. The results highlight design opportunities that have an impact on both research and real-world implementation. These prospects encompass deep learning, recommender systems, sound identification, and chatbot engagement with additional fashion applications.

III. PROPOSED METHOD

This section demonstrates the robust architecture of the system. All the modules developed with multiple DL and NLP techniques. The schematic system architecture of our proposed technique is depicted in Figure. 1. Initially, the user initiates a query, and in the event of a congruence with the database, the system proceeds to the relevant module; otherwise, it furnishes feedback to the user. Subsequent to capturing the user's response, as delineated in Fig 1, the system engages in ASR to convert the user's textual input into an audible format. The textual data undergoes encoding through the utilization of the sentence transformer model, specifically "paraphrase-mpnet-base-v2," was employed. The

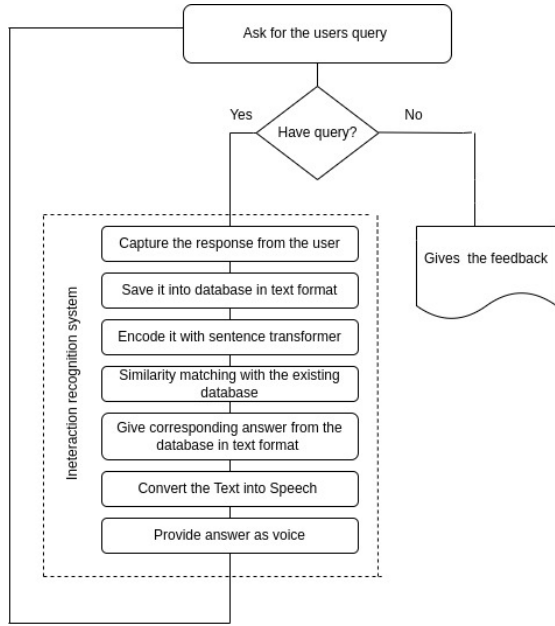


Figure 1. System architecture of the voice based chatbot

system, thereafter, conducts a similarity assessment of the encoded data using cosine similarity metrics against the extant database. Following the similarity matching process, the system furnishes the most pertinent response from the database. These intricate procedures are facilitated by the Interaction Recognition System.

A. Dataset Description

Data has been collected from the First Security Islami Bank, involving multiple visits spanning from November 21 to November 23, 2023. During these visits, interactions were conducted with the bank manager, two additional staff members, and several customers to elicit queries and corresponding answers. The resultant dataset comprises 152 entries in CSV format [14].

B. Interaction Recognition System

The development of the interactive agent is enhanced by the integration of ASR, Interactive Agent, and TTS synthesis modules, as illustrated in Fig 2. The system acquires user vocal input through a microphone, wherein ASR proficiently transcribes the spoken content into Bengali text format utilizing gTTS. Subsequently, the transcribed text undergoes encoding within the database, employing the "paraphrase-mpnet-base-v2" sentence transformer model. Within the database, cosine similarity is utilized to compare the embedded data with pre-established questions, facilitating the retrieval of corresponding answers. Finally, the TTS system translates the textual response of the interactive agent into synthesized speech.

C. Interactive Agent in Bengali

The primary objective of the Interaction Recognition system is to receive textual inquiries from the ASR system and subse-

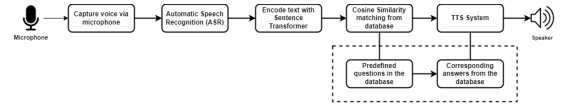


Figure 2. Workflow of the proposed method

quently provide the most pertinent response from a knowledge-based database. This Interactive Agent is specifically tailored for the domain of banking customer care services. The database encompasses queries and corresponding responses pertaining to banking matters, all articulated in the Bengali language. The sentence embedding process is facilitated by the utilization of a sentence transformers model, specifically "paraphrase-mpnet-base-v2". As illustrated in Figure 3, the bespoke Interactive Agent can be delineated into five sequential stages.

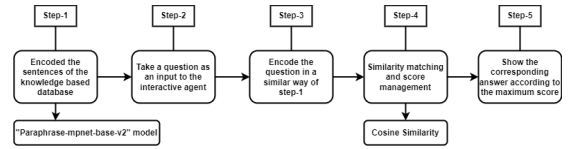


Figure 3. Steps of Interactive Agent

Step 1: Utilizing the Sentence Transformer, we undertake the encoding of all sentences within the knowledge-based database. The database is stratified into two distinct segments, namely the interrogative component and the corresponding response section. Employing the "paraphrase-mpnet-base-v2" model for sentence embedding, we systematically convert each normalized question and its corresponding answer into a vectorized format.

Step 2: Contemplate the ensuing inquiry presented to the Interactive Agent: The procedural sequence commences by receiving a query articulated by a system user through ASR, specifically designed for voice input pertaining to inquiries related to banking. The transmutation of the received textual information into the Bengali language is facilitated through the utilization of the 'paraphrase-mpnet-base-v2' model, after which the processed content is subsequently dispatched to the Interactive Agent.

Step 3: The imperative to encode the sentence provided by users is deemed essential. Employing analogous methodologies as delineated in the initial step, sentence embedding procedures were implemented for the aforementioned purpose.

Step 4: Facilitating Similarity Matching and Score Measurement entails the strategic enumeration of recurrent terms within the interrogative segment of the database juxtaposed with user input, serving as a viable approach in crafting a closed-domain Interactive Agent. Nevertheless, challenges may manifest when confronted with an expansive corpus, as the prevalence of shared terms escalates proportionally with the magnitude of the prepared textual content. To mitigate this, the cosine similarity metric is employed to ascertain optimal scores in evaluating the congruence between the user's inquiry

and database-interrogative entries.

Step 5: Upon receipt of a user query, the system generates an appropriate response by employing cosine similarity calculations to ascertain the highest attainable score between the formulated questions and user input. The resultant output from the Interactive Agent is presented exclusively in textual form, serving as the input for subsequent text-to-speech synthesis.

D. Text-to-Speech Synthesis

To cultivate a voice-enabled conversational agent capable of addressing diverse inquiries pertaining to banking, it is imperative to generate audible responses. Initially, the translation of textual input into speech is facilitated through the integration of the 'gTTS' module. It enhance user experience and accessibility. It is fast, lightweight and easy to integrate.

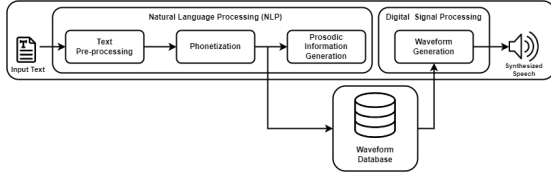


Figure 4. Text-to-Speech (TTS) synthesis

E. Web Application of the 'voice based conversational chat-bot'

The interactive agent is equipped with a web application developed using Flask, a compact and lightweight Python web framework. Illustrated in Figure 8, the user interface of the web application is denoted as "AI Reception". Upon launching the application, users are greeted with a welcome message and a microphone button. To initiate queries, users must press the microphone button and observe a brief 1.5-second delay. Subsequently, users can articulate their inquiries, and the interactive agent will respond with the most probable answer in spoken form. The system retains the spoken response for auditory playback prior to the user posing another query.

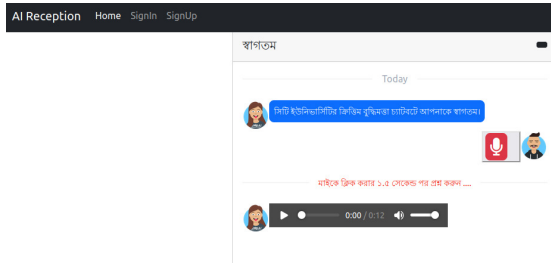


Figure 5. Web application of the voice based conversational chatbot

IV. RESULTS AND DISCUSSION

This interactive agent facilitates the conversion of human speech input into textual data and vice versa, employing the gTTS technology. The system employs sentence transformers to encode queries, subsequently employing cosine similarity to identify the optimal response from a database.

Figure 5 illustrates a potential outcome resulting from an appropriately formulated query. Furthermore, the provided



Figure 6. Input and Output of Right Query and Answer

inquiry is preserved within the input.txt file in text format, while the corresponding verbal output is stored as an audio file and user can listen the audio from the web application. Subsequent to each query, these files undergo replacement with updated text and audio content. Additionally, the user sometime poses erroneous inquiries that lack relevance to the predetermined database. Consequently, the system provides feedback to apprise the user of such discrepancies.



Figure 7. Feedback to the wrong query

This interactive conversational agent facilitates the conversion of human speech input into textual data and vice versa through the utilization of gTTS technology. The system's efficacy has been validated through rigorous testing with diverse queries, yielding similarity scores for each inquiry. Recorded queries are persistently stored as text files, while the corresponding responses are saved in audio format. In instances where vocal inputs are not accurately recognized, the system provides constructive feedback to users in auditory form. Additionally, the system offers feedback in speech format for inappropriate queries. Subsequently, a web application has been developed to enhance the system's dynamism and user accessibility.

V. CONCLUSION

The proposed system delineates an advanced AI-driven interactive agent designed for the banking reception milieu, aiming to optimize user satisfaction through the provision of voice-enabled question-answering capabilities in the Bengali language. The system encompasses three pivotal modules: ASR convert the user's textual input into an audible format; an Interactive Agent empowered by the "paraphrase-mpnet-base-v2," enhancing its linguistic comprehension and responsiveness to user inquiries through Bengali text encoding; and TTS Synthesis module for the instantaneous generation of human-like responses. The cohesive integration of these modules culminates in the development of a responsive interactive agent, representing a noteworthy advancement in contemporary customer care systems for banking reception, adept at efficiently managing and addressing customer queries in the Bengali language.

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] P. Klaus and J. Zaichkowsky, "AI voice bots: a services marketing research agenda," *J. Serv. Mark.*, vol. 34, no. 3, pp. 389–398, 2020.
- [9] R. Burri, "Improving user trust towards conversational chatbot interfaces with voice output," 2018.
- [10] S. Quarteroni and S. Manandhar, "A chatbot-based interactive question answering system," *Decalog*, vol. 83, 2007.
- [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. 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.
- [13] 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.
- [14] M. N. Huda Nahid Khandaker, M. M. Hossain, and M. T. H. Tusar, "Bengali Chatbot QA Data For Banking Domain." *Kaggle*, 2024, doi: 10.34740/kaggle/ds/4839725