VOICE ASSISTANT BASED ON GPT AND BING FOR HEALTHCARE

Submitted in partial fulfillment of the requirements for the degree of

Bachelor of Technology

In

Computer Science and Engineering

By

Saraswathi Bavadharini S

20BCE0339

Under the guidance of

Prof. Manoov R

School of Computer Science and Engineering

VIT, Vellore.

February

2024

ABSTRACT

This project introduces a novel voice assistant system meticulously crafted to cater to the intricacies of healthcare applications. It harnesses the prowess of advanced natural language processing models, notably GPT-3.5 Turbo and Bing, to deliver tailored support for healthcare-related inquiries and tasks. The system is distinguished by its utilization of specialized wake words, "Bing" and "GPT," enabling users to seamlessly activate specific services tailored to their needs within the healthcare domain. By prioritizing accessibility and efficiency, this voice assistant aims to revolutionize healthcare interactions, offering intuitive assistance for both healthcare professionals and patients. Through a combination of cutting-edge technology and domain-specific functionality, it endeavors to elevate the healthcare experience to new heights.

INTRODUCTION

Voice assistants have emerged as indispensable tools in various domains, offering convenience and efficiency in everyday tasks. However, existing voice assistant solutions often lack the specificity and functionality required to address the unique challenges of the healthcare sector. In response to this gap, this project introduces a dedicated voice assistant system tailored explicitly for healthcare applications. By integrating advanced natural language processing models like GPT-3.5 Turbo and Bing, the system aims to provide tailored assistance for healthcare professionals and patients, facilitating efficient information retrieval, task management, and communication within healthcare environments. Through a combination of cutting-edge technology and domain-specific functionality, this voice assistant endeavors to revolutionize healthcare interactions, offering intuitive support that enhances efficiency and improves the overall healthcare experience for users.

OBJECTIVE

The primary aim of this project is to develop a sophisticated voice assistant system meticulously designed to meet the unique demands of the healthcare sector. By leveraging the capabilities of state-of-the-art natural language processing models such as GPT-3.5 Turbo and Bing, the system seeks to provide comprehensive and contextually relevant support for a wide range of healthcare-related inquiries and tasks. Through intuitive voice-based interactions and the implementation of specialized wake words, the objective is to enhance accessibility, streamline workflows, and improve overall efficiency in healthcare delivery and patient care. Ultimately, the goal is to empower healthcare professionals and patients alike with a versatile and efficient tool that facilitates seamless communication, information retrieval, and task management within healthcare environments.

PROBLEM STATEMENT

Despite the growing demand for voice assistant solutions in healthcare, existing systems often struggle to meet the specialized needs of this domain. Challenges include the integration of advanced natural language processing capabilities, the development of tailored functionality for healthcare-specific tasks, and the need for seamless communication within healthcare environments. Consequently, there is a pressing need to develop a dedicated voice assistant solution optimized for healthcare applications. Such a solution would leverage cutting-edge models like GPT-3.5 Turbo and Bing to deliver accurate, contextually relevant assistance, thereby enhancing efficiency and improving the overall healthcare experience for users. By addressing these challenges, this project seeks to fill a crucial gap in the market and provide healthcare professionals and patients with a versatile tool that meets their unique needs and requirements.

LITERATURE SURVEY

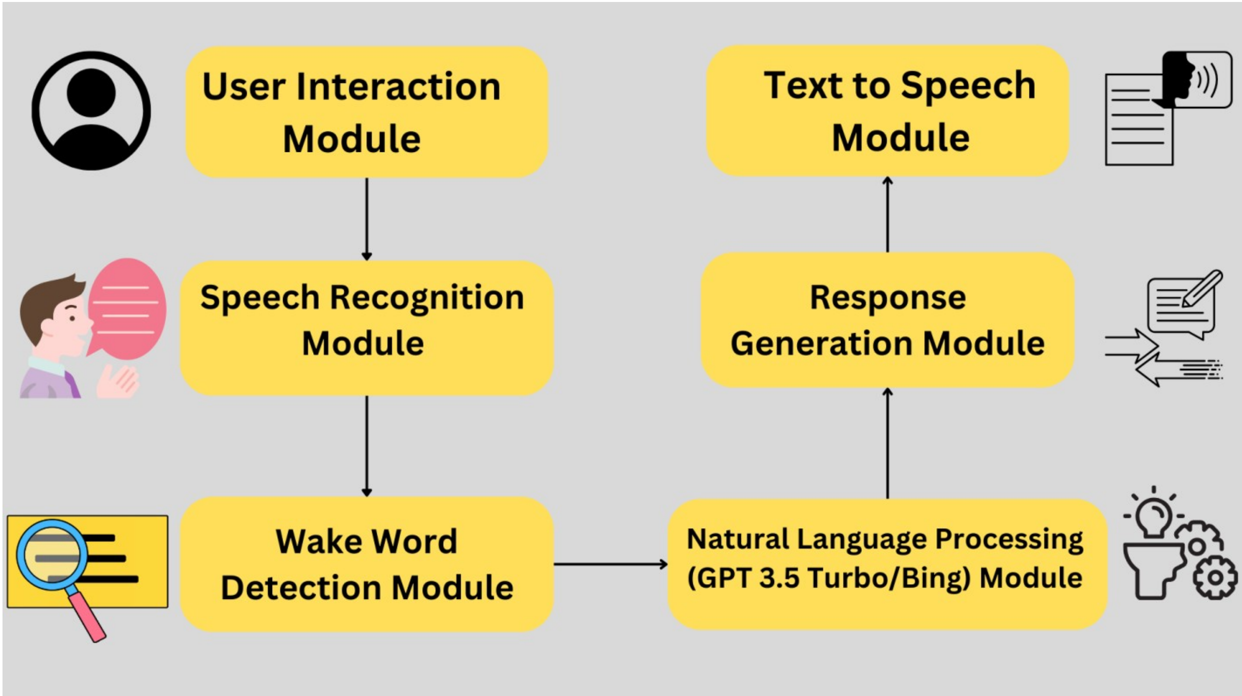
Ref.	Paper Title	Author and Year	Research Question	Methodology/Approach	Limitation
1	Benefits, Limits, and Risks of GPT-4 as an AI Chatbot for Medicine	Lee *et al.* (2023)	Using GPT 4 and similar Generative AI tools such as Google LaMDA and GPT 3.5 in Medical conversational ChatBOTs	ChatBOTs use the GPT 4 LLM to retrieve ansers for user queries from web and this model has been tested out and found to have an accuracy of over 90%	Authenticity of Data obtained from the web by GPT models.
2	Speech emotion recognition using machine learning - A systematic review	Madanian *et al*. (2023)	Properties, methodology and working of SER model and analysing its efficiency .	Training a speech recognition (SR) system, including language corpus, nursing activities, clinical conversations, and accents. It compared documentation time and error rates between SR-generated records and keyboard entry,	The paper may overlook non-ML approaches and interdisciplinary perspectives in SER, and while it discusses challenges and solutions, it may not encompass all potential obstacles

					or emerging trends.
3	Development of the Speech-to-Text Chatbot Interface Based on Google API	Shakhovska *et al.* (2019)	Utilizing the Google Speech-to-Text API.data from social networks to focused on remote and local storage processes.	The proposed method involves employing prefix functions and hashing algorithms for keyword searching and verb ending identification in chatbot conversations	The study may potentially overlook alternative methods and their effectiveness in real-world applications.
4	Machine learning-based speech recognition system for nursing documentation – A pilot study	Lee *et al.* (2023)	Machine learning-based speech recognition (SR) system's effectiveness in reducing nursing documentation workload in a psychiatry ward.	The study collected language corpus, nursing activities, clinical conversations, and accent data for SR system training in four sessions and achieved model had an accuracy score of 87.06% to 95.07% across sessions.	The study's findings are based on a pilot implementation in a psychiatry ward, potentially limiting generalizability to other nursing specialties or healthcare settings.
5	Intelligent speech technologies for transcription, disease diagnosis, and medical equipment	Zhang *et al.* (2023)	To explore the application and potential of intelligent speech technology (IST) in addressing medical resource	The paper introduces IST's procedure and system architecture, reviews its applications in smart hospitals, and presents a case study on stroke patient care. Additionally, it proposes a novel medical voice	Challenges include noise interference and pronunciation differences, which may hinder the widespread application of IST in hospitals.

	interactive control in smart hospitals: A review		shortages and improving healthcare efficiency amid challenges like noise interference and pronunciation differences.	analysis system architecture.	
6	The Capability of ChatGPT in Predicting and Explaining Common Drug-Drug Interactions	Juhi A *et al.* (2023)	To assess the effectiveness of ChatGPT in predicting and explaining common drug-drug interactions (DDIs)	Utilized 40 DDI lists from literature to converse with ChatGPT using two-stage questions, assessing responses' correctness with pharmacologists' consensus.	ChatGPT provided incomplete guidance at times, necessitating further improvement for patient use regarding DDI awareness.
7	Deep Cross-Corpus Speech Emotion Recognition: Recent Advances and Perspectives	Zhang *et al.* (2021)	To comprehensively survey the state-of-the-art techniques in cross-corpus speech emotion recognition (SER), particularly focusing on deep learning methods	The paper reviews existing literature on speech emotion databases, traditional methods for cross-corpus SER, recent advances in deep learning techniques, and discusses challenges and future directions in the field.	Challenges such as natural data scarcity, multimodal integration, and limitations of deep learning techniques, potentially affecting the comprehensiveness of its findings are discussed.

			associated with supervised, unsupervised, and semi-supervised learning.		
8			Propose Contrastive Language-Audio Pretraining (CLAP) for joint audio-text representation learning, enabling Zero-Shot inference across 26 downstream tasks, surpassing state-of-the-art models for general-purpose audio representations.	Utilize two innovative encoders for audio and text, trained with Contrastive Learning to create multimodal representations. Train audio encoder (HTSAT-22) on 22 tasks and adapt GPT2 for text encoding, enabling joint learning of representations in a multimodal space for Zero-Shot inference.	Limited evaluation on tasks with true Zero-Shot setup. Increased training pair diversity affects performance variably across domains.

SYSTEM MODEL WITH DESCRIPTION



- User Interaction Layer: This is the interface through which users interact with the voice assistant system.
- Speech Recognition Module: Responsible for converting spoken words into text format.
- Wake Word Detection Module: Identifies specific wake words ("Bing" and "Gpt") to activate the system. Using pattern recognition and machine learning techniques, it quickly identifies these triggers, reducing latency and ensuring fast system activation.
- Response Generation Module: Generates appropriate responses based on the processed user queries.
- Text-to-Speech Module: Converts the generated responses into spoken words for the user to hear.

REFERENCES

1. Juhi A, Pipil N, Santra S, Mondal S, Behera JK, Mondal H. The Capability of ChatGPT in Predicting and Explaining Common Drug-Drug Interactions. Cureus. 2023 Mar 17;15(3):e36272. doi: 10.7759/cureus.36272. PMID: 37073184; PMCID: PMC10105894.
2. Lee P, Bubeck S, Petro J. Benefits, Limits, and Risks of GPT-4 as an AI Chatbot for Medicine. N Engl J Med. 2023 Mar 30;388(13):1233-1239. doi: 10.1056/NEJMsr2214184. PMID: 36988602.
3. Chi-Chun Lee, Emily Mower, Carlos Busso, Sungbok Lee, Shrikanth Narayanan,Emotion recognition using a hierarchical binary decision tree approach,Speech Communication,Volume 53, Issues 9–10,2011,Pages 1162-1171,ISSN 0167-6393.
4. Imani, Maryam & Montazer, Gholam Ali. (2019). A survey of emotion recognition methods with emphasis on E-Learning environments. Journal of Network and Computer Applications. 147. 102423. 10.1016/j.jnca.2019.102423.
5. Anagnostopoulos, Christos-Nikolaos & Iliou, Theodoros & Giannoukos, Ioannis. (2012). Features and classifiers for emotion recognition from speech: a survey from 2000 to 2011. Artificial Intelligence Review. 43. 10.1007/s10462-012-9368-5.
6. Liu, Zhen-Tao & Wu, Min & Cao, Weihua & Mao, Jun-Wei & Xu, Jian-Ping & Tan, Guan-Zheng. (2017). Speech emotion recognition based on feature selection and extreme learning machine decision tree. Neurocomputing. 273. 10.1016/j.neucom.2017.07.050.
7. Ghosh, Sayan, Eugene Laksana, Louis-Philippe Morency, and Stefan Scherer. "Representation learning for speech emotion recognition." In Interspeech, pp. 3603-3607. 2016.
8. Jahangir, Rashid, Ying Wah Teh, Faiqa Hanif, and Ghulam Mujtaba. "Deep learning approaches for speech emotion recognition: State of the art and research challenges." Multimedia Tools and Applications (2021): 1-68.

9. Fahad, Md Shah, Ashish Ranjan, Jainath Yadav, and Akshay Deepak. "A survey of speech emotion recognition in natural environment." Digital signal processing 110 (2021): 102951.
10. Lee, Tso-Ying, Chin-Ching Li, Kuei-Ru Chou, Min-Huey Chung, Shu-Tai Hsiao, Shu-Liu Guo, Lung-Yun Hung, and Hao-Ting Wu. "Machine learning-based speech recognition system for nursing documentation–A pilot study." International Journal of Medical Informatics 178 (2023): 105213.
11. Singh, Aashdeep, P. Murugeswari, SD Prabu Ragavendiran, Amanpreet Kaur, Gurpreet Singh, and Sathiyamoorthy Margabandu. "AI-based Chatbot for Physically Challenged People." In 2022 International Conference on Edge Computing and Applications (ICECAA), pp. 1039-1044. IEEE, 2022.
12. Shakhovska, Nataliya, Oleh Basystiuk, and Khrystyna Shakhovska. "Development of the Speech-to-Text Chatbot Interface Based on Google API." In MoMLeT, pp. 212-221. 2019.
13. Madanian, Samaneh, Talen Chen, Olayinka Adeleye, John Michael Templeton, Christian Poellabauer, Dave Parry, and Sandra L. Schneider. "Speech emotion recognition using machine learning—A systematic review." Intelligent systems with applications (2023): 200266.
14. Zhang, Jun, Jingyue Wu, Yiyi Qiu, Aiguo Song, Weifeng Li, Xin Li, and Yecheng Liu. "Intelligent speech technologies for transcription, disease diagnosis, and medical equipment interactive control in smart hospitals: A review." Computers in Biology and Medicine (2023): 106517.
15. Choudhary, Nurendra, and Chandan K. Reddy. "Complex Logical Reasoning over Knowledge Graphs using Large Language Models." arXiv preprint arXiv:2305.01157 (2023).
16. Elizalde, Benjamin, Soham Deshmukh, and Huaming Wang. "Natural language supervision for general-purpose audio representations." arXiv preprint arXiv:2309.05767 (2023).