ISSN: 2454-9940

# INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN; 2454-9940 www.ijsem.org
Vol 17, Issuse.1 March 2023

# AI CHATBOT WITH VOICE ASSISTED ANSWER

**[1] T.Rajesh, [2] V.Sravanthi,[3] M.Srujana, [4] V.JAYASREE**

[1,2,3] Assistant Professors,Department of Computer Science and Engineering, Kasireddy Narayanreddy College Of Engineering And Research, Abdullapur (V), Abdullapurmet(M), Rangareddy (D), Hyderabad - 501 505

[4]student,Department of Computer Science and Engineering, Kasireddy Narayanreddy College Of Engineering And Research, Abdullapur (V), Abdullapurmet(M), Rangareddy (D), Hyderabad - 501 505

## ABSTRACT

In today's increasingly digital world, chatbots powered by artificial intelligence (AI) have become indispensable for improving user interactions in many fields, such as healthcare, education, entertainment, and customer support. Building an artificial intelligence chatbot with a voice-assistant answering system is the main goal of this project. The goal is to make the interaction between humans and machines more natural and productive. To provide users with a natural conversational experience, the chatbot uses state-of-the-art Speech Recognition and Natural Language Processing (NLP) technology. The system is able to comprehend both text-based input and voiced questions because to its voice interaction features, which allow users to access the interface without using their hands.Adaptable to a wide range of mediums, including websites, mobile apps, and voice-activated devices, the voice-assisted chatbot may reply to user enquiries with either text or speech outputs. It learns from its users' actions and preferences over time to provide better replies and meet their varied demands. A customised experience is guaranteed by the system's multilingual support and its capacity to manage context-aware interactions. By making information more accessible and enhancing user involvement, this project hopes to increase user satisfaction via the provision of faster and more accurate responses in aural and visual forms.

# I.INTRODUCTION

Customer service, healthcare, education, and e-commerce are just a few of the many sectors that have found AI-powered chatbots to be indispensable tools in the quest for more accessible, efficient, and user-friendly solutions. Though successful, chatbots have traditionally only been able to connect with users via text. This limits their ability to meet the varied demands of their users since it lacks the personalisation and dynamic engagement that is required. A rising trend is the integration of voice-based interfaces, which turn static chatbots into more engaging, hands-free systems, thereby overcoming these restrictions.

The goal of this project is to create an artificial intelligence chatbot that can answer user questions with the use of voice recognition and other features that integrate natural language processing (NLP), machine learning (ML), and speech recognition. The incorporation of voice interaction is the main novelty of this project. It enables users to communicate with the chatbot using both text and speech. Users with impairments or a preference for voice communication will find the system more accessible thanks to this integration, and the user experience will be enhanced by quicker and more intuitive replies.

With its real-time query understanding and processing capabilities, the suggested system can provide pertinent responses in text and audio forms. The chatbot provides a tailored, context-aware experience that changes with the user's preferences as time goes on, thanks to its extensive speech recognition and response capabilities. There will be more efficient customer service, happier users, and a more inclusive digital experience overall thanks to this voice-assisted chatbot. In addition, it can be used on many platforms, including as websites, mobile apps, and gadgets that can be controlled by speech, making it a flexible tool for companies and individuals alike. The end objective of this project is to build an AI-powered system that can answer questions and have genuine conversations with people, making technology more accessible and interactive for everyone.

# II.LITERATURE REVIEW

Due to the rising need for efficient and personalised user interactions across several sectors, research and development efforts have mostly focused on integrating AI with chatbot technologies in recent years. Accessibility, user engagement, and the

capacity to provide a more natural interaction are all areas where traditional chatbots fall short. These bots depend entirely on text input and output. In response to these difficulties, researchers have begun to use voice-assisted chatbots, which integrate NLP and speech recognition to allow users and AI systems to communicate in a more natural and conversational way. In order to shed light on the strengths, weaknesses, opportunities, and threats associated with developing an AI chatbot capable of voice-assisted response, this literature review surveys previous research and technological developments in the fields of artificial intelligence (AI), voice interaction, natural language processing (NLP), and speech recognition.

## 1. AI Chatbots and Natural Language Processing (NLP)

Chatbots powered by artificial intelligence make use of Natural Language Processing (NLP) to simulate human conversational understanding, interpretation, and response. A branch of artificial intelligence, natural language processing (NLP) aims to teach computers to read, write, and comprehend spoken and written language. Tokenisation, part-of-speech tagging, named entity identification, and sentiment analysis are some of the natural language processing (NLP) activities that chatbots utilise to understand user enquiries and come up with relevant answers (Joulin et al., 2017). Approaches to these challenges in natural language processing (NLP) range from rule-based models to retrieval-based models and even generative models. According to Devlin et al. (2018), AI chatbots' capacity to comprehend context and provide more relevant replies has been greatly enhanced by transformer-based designs like BERT and GPT. Artificial intelligence chatbots have become more useful in industries including education, healthcare, and customer service as a result of these developments.

Traditional text-based chatbots still have a ways to go before they can provide a truly immersive experience for users, particularly those with visual impairments or low literacy, despite recent developments. The incorporation of voice interaction and speech recognition into AI chatbot systems has thus become an increasingly popular area of research.

## 2. Voice-Assisted Chatbots

By allowing consumers to communicate with AI via spoken language, voice-assisted chatbots constitute a notable

improvement over text-based systems. These chatbots shine in situations when people would rather not use both hands, including while driving, cooking, or dealing with physical limitations that hinder their typing ability. The possibility of incorporating speech recognition into interactive systems has been shown by voice-based AI systems that have been popularised by voice assistants like Apple Siri, Google Assistant, and Amazon Alexa. Chowdhury et al. (2020) states that voice assistants have gained significant traction because of their capacity to understand and respond to natural language requests across a variety of devices and platforms, including smartphones, smart speakers, and smart home gadgets.

Technologies that convert text to speech (TTS) and speech to text (STT) are useful for voice-assisted chatbots. Using natural language processing (NLP) methods, STT systems transcribe spoken language into text. On the other side, text-to-speech algorithms make sure the chatbot can talk to the user by turning text into voice that sounds natural. By allowing for verbal engagements without the need to utilise one's hands, the integration of these technologies improves the user experience. Using a mix of STT and TTS, Vasilenko et al. (2019) shown that chatbot systems are far more accessible, especially for those who have trouble seeing or who prefer audio to text.

While voice-assisted chatbots do make information more accessible, there are still significant obstacles, such as how accurate speech recognition is, how well they comprehend context, and how natural their synthesised speech sounds. An problem that might arise in the user experience is the difficulty of effectively transcribing spoken input in loud locations or with users who have strong accents, as pointed out by Garcia et al. (2018). Further, there is still a long way to go in terms of research into creating convincing speech syntheses that seem natural and interesting.

## 3. Speech Recognition and Machine Learning

Transcribing audio impulses into text is the essence of speech recognition technology, which allows for voice interaction. To enhance accuracy and flexibility, modern voice recognition systems use complex machine learning techniques, such RNNs and deep learning, in products like Microsoft Azure's voice Service and Google's Speech-to-Text API. In their 2019 study, Liu et al. demonstrated that LSTM networks and other deep learning approaches have greatly improved speech recognition

models by tackling issues with recognising context and temporal dependencies.

Both natural language processing (NLP) and voice recognition are greatly enhanced by machine learning. The system's accuracy increases with time as it learns to identify human speech patterns and subtleties via more input. By improving context retention and allowing parallel processing of huge datasets, transformer models proposed by Vaswani et al. (2017) have transformed natural language comprehension and audio recognition. These improvements are especially helpful for building voice-assisted chatbots since they let the system understand user questions better and provide more personalised answers.

Maintaining real-time performance, dealing with many languages and dialects, and attaining high identification accuracy in varied situations are still hurdles, despite recent developments. Hybrid models, which combine rule-based and machine learning techniques, may enhance the flexibility and resilience of speech recognition systems in different acoustic environments, according to researchers such as Nassir et al. (2020).

## 4. User Experience and Interaction Design

The UX and interface design of voice-assisted chatbots are crucial to their success. An efficient voice interface can understand the user's context and respond with clarity and brevity. The importance of a realistic conversational flow and little friction in voice interfaces was highlighted by Schmidt et al. (2018). When designing for user satisfaction, it's important to consider things like reaction speed, vocal intonation, clarity, and emotional recognition. According to research by Sundararajan et al. (2020), chatbots that have emotional intelligence may greatly enhance user engagement and happiness by recognising user emotions and responding appropriately.

In order to create a voice-assisted chatbot that can be used all over the world, it is essential that it can understand and process many languages and dialects. The difficulties of training models to accurately handle many languages and accents were brought to light by Gusmão et al. (2019), who investigated multilingual voice assistants.

## 5. Applications of AI Chatbots with Voice Assistance

Many different types of businesses are quickly embracing AI chatbots that have voice-assisted responding systems. In the field of customer service, they are able to

decrease wait times and increase customer satisfaction by providing quicker and more personalised replies. A hands-free alternative to conventional healthcare systems, voice-assisted chatbots may help patients with appointment booking, prescription reminders, and health enquiries. Tutoring students, answering their queries, and facilitating learning activities are just a few ways in which voice-assisted AI chatbots are making a splash in the education industry. An online store's customer service, product searches, and order placements may all be improved with the use of a voice-powered assistant. Because speech interactions are becoming more commonplace and natural, these technologies are finding their way into more and more mobile applications, smart speakers, and wearables. Businesses now have more ways than ever to connect with consumers and improve service thanks to voice-assisted AI chatbots, which combine natural language processing (NLP), speech recognition (SR), and machine learning (ML).

**6. Challenges and Future Directions**

Accuracy, contextual comprehension, and personalisation continue to be hurdles for artificial intelligence chatbots and speech recognition technology, despite their progress. Fully conversational AI that can grasp context, answer complicated enquiries, and participate in multi-turn discussions is still a big research problem, as pointed out by Rosenberg et al. (2020). More importantly for the future of speech synthesis, there is a need to enhance it so that it can generate realistic, emotionally intelligent speech.

Protecting users' personal information and financial data while using voice-assisted AI systems might be a future area of research. For these systems to gain traction, it is essential to guarantee the safe processing of speech data while also protecting users' privacy.

## III.PROPOSED MODEL

### A. Study Data

A number of datasets are used to train and evaluate the various machine learning models that drive the functionality of the AI Chatbot with Voice-Assisted Answering project. Ensuring that the chatbot is trained to comprehend and handle customer enquiries requires text-based conversational data. To allow the chatbot to manage varied conversational circumstances and reply correctly, datasets such as Persona-Chat, DailyDialog, and the Cornell Movie

Dialogues are used. The process of transcribing spoken language into writing requires both text data and voice recognition data. To train the speech recognition model to correctly transcribe user voice inputs, taking into consideration diverse accents and speech patterns, datasets like LibriSpeech, CommonVoice, and TED-LIUM are used. To make sure the chatbot sounds genuine and responds naturally, Text-to-Speech (TTS) data from sources like VCTK and LJSpeech is used to train the system. The chatbot's conversational skills are enhanced by using multi-modal datasets like MOSI and AVE, which let it grasp both the text and the emotional tone of voice inputs. The chatbot's performance in real-world settings is measured using evaluation and testing data, which includes unique chatbot evaluation datasets and voice recognition test sets. Also, after deployment, you may gather data on user interactions like feedback and discussion logs to make sure the system is always running better. The project's overarching goal is to construct a powerful voice-assisted chatbot that can improve the user experience by providing replies that are accurate, context-aware, and emotionally intelligent by integrating various datasets.

**B) System Architecture**

A variety of complicated activities may be handled by the AI chatbot with voice-assistant answering system architecture, which guarantees a smooth user-bot interaction. To provide a powerful and efficient conversational AI experience, it combines voice recognition, NLU, dialogue management, response generation, and text-to-speech (TTS) technologies. A number of interdependent modules make up the system, and they all work together to provide the desired result.

The UI, or user interface:

The user interface (UI) is the part of a chatbot that the user interacts with on the front end. It lets the user choice between using their fingers or their voice to enter data. The user talks into a microphone or a smart device (such as a mobile phone or a smart speaker) while using voice input. Web sites, smartphone applications, and social media platforms often include chatbot interfaces where users may enter their queries. The UI is designed to be easy to use and understand, so users may transition between text and speech interactions without any hiccups.

The Speech Recognition Module is in charge of transcribing audio files into text when the user chooses to utilise voice input. It uses state-of-the-art models, such as DeepSpeech or Google's Speech-to-Text API, to process the original audio

input. Regardless of the user's demography, the system will function reliably since it can handle several languages and dialects. The system employs acoustic and linguistic models to correctly comprehend human speech after using noise reduction methods to remove unnecessary background noise. The next part of the system processes the speech once it is transcribed into text.

Understanding Natural Language (NLU): The NLU module analyses the identified text after the speech-to-text conversion. Here, the chatbot's interpretation of the user's input is crucial. The NLU component uses natural language processing (NLP) methods like tokenisation, lemmatisation, and part-of-speech tagging to deconstruct the text into manageable chunks and extract relevant information like named entities, intentions, and connections between query components. By analysing the input, it deduces the user's request and its meaning. When a user asks, "What's the weather going to be like in New York tomorrow?" the Natural Language Understanding (NLU) module recognises "weather" as the intent and "New York" and "tomorrow" as entities.

To guarantee proper understanding of even subtle or complicated questions, this stage might make use of advanced models such as BERT or Rasa NLU. Further, the NLU module is responsible for handling context, which means it may recall past encounters to keep things consistent and provide answers that are more appropriate to the current situation. The Dialogue Management System (DMS) takes over conversation flow management once intent and entities have been retrieved. The DMS considers the user's choices, the current context of the discussion, and any potential follow-up enquiries to decide how the system should answer. In other words, it determines whether the user's inquiry doesn't need a tailored response or if it can be replied straight using established templates.

Third, the system may look up the answer in a database of frequently asked questions (FAQs) or a knowledge base if it contains prepared solutions. But when the question calls for a tailored response, the DMS talks to the Response Generation Module. Natural and human-like text answers may be generated by this module using complex language models like GPT-3 or BERT-based transformers. The chatbot may learn from repeated interactions, picking up on nuanced changes in tone or user actions to provide personalised replies.

Transcribing Text into Speech (TTS):

If the user has selected voice-assisted interaction, the chatbot will next

transform the text-based answer into speech. The Text-to-Speech (TTS) Module converts the written text into audio that sounds like human speech. The use of deep learning models in modern TTS systems like WaveNet or Google Cloud Text-to-Speech allows for the production of human-like voice answers that fluctuate in pitch, emotion, and tone, resulting in a more engaging and lifelike discussion. The system may also change the loudness and speech tempo according to the user's preferences, making it more personalised. In order to make the chatbot's speech answers seem more natural and less artificial, the TTS module integrates prosody elements such as pauses, emphasis, and intonation. To further enhance the user experience, the chatbot might include a variety of voice choices, such as male, female, and other accents. The user's output device, such a smartphone speaker or smart speaker, plays back the audio once the speech is generated by the system.

Criticism and Ongoing Development:

Ability to continually develop based on user input is a fundamental characteristic of current AI chatbots. During conversations, the AI chatbot with voice-assisted answering collects and feeds back both implicit (such as the amount of time it takes for a user to respond or the length of a message) and explicit (such as a thumbs up or down) user input. With this input, we can make the model even better, which in turn makes voice recognition and answer generation more accurate and, ultimately, makes the user experience better. To be current and successful, the system may update its language models frequently to react to changing language trends and user preferences.

Customisation and User Context:

The system design also includes a Contextual Understanding module to improve the user experience. This section keeps tabs on current discussions, recalls past questions, and uses past data to provide better tailored answers. As an example, to enhance the user experience, the system might enquire about the user's preferred frequency of weather updates (daily, weekly, etc.) in response to repeated weather-related queries. Personalisation of calendar events, reminders, and preferences is also possible via system integration with user accounts (e.g., cloud services, smart home devices).

## IV.CONCLUSION

The AI Chatbot with speech-Assisted Answering system is a huge step forward in AI since it combines speech recognition, NLU, and real-time answer

creation in a way that's easy for users to comprehend and utilise. The goal of this system's design was to provide users with more user-friendly interactions by using cutting-edge technology such as deep learning models for voice recognition and natural language processing (NLP). Whether users want to communicate with the system via hands-free voice commands or direct text interactions, the system's versatility allows it to accommodate both types of inputs.

By combining speech-to-text and text-to-speech technologies, accessibility is improved, making technology more inclusive for people with various requirements, including physical limitations. In addition, the system may adapt to user behaviour by learning from experience, incorporating feedback, and personalising it. As a result, it becomes better at offering relevant and accurate responses over time.

The suggested system design does a good job of addressing the problems with voice-assisted interactions, but it may be much better. Incorporating more sensory inputs into multimodal engagement, making speech recognition more accurate in loud surroundings, and better comprehending context for more personalised discussions are all on the list. The increasing demand for AI-driven voice-based interactions in sectors like education, customer service, and healthcare suggests that future research should concentrate on improving the system's performance to keep up with the increasing number of user queries.

## V.REFERENCES


1. J. Y. Lee, S. Kim, and H. Lee, "A conversational AI chatbot for education," *Proceedings of the International Conference on Artificial Intelligence*, vol. 23, pp. 234-238, 2021.

2. M. Zhang, X. Liu, and H. Zhang, "Speech recognition for chatbot applications," *IEEE Transactions on Signal Processing*, vol. 67, no. 10, pp. 2601-2609, 2020.

3. K. S. Sharma and S. K. Gupta, "Natural language understanding for conversational agents: A review," *Journal of Computer Science & Technology*, vol. 36, pp. 1024-1039, 2021.

4. S. B. Lee and Y. S. Lim, "Deep learning-based chatbot with voice interaction: Implementation and challenges," *IEEE Access*, vol. 9, pp. 16023-16033, 2021.

5. T. K. Roy and M. R. K. Saha, "Improving speech-to-text accuracy with

deep learning techniques," *Speech Communication*, vol. 115, pp. 71-79, 2019.

6. D. S. Bhattacharya, "A review on speech recognition systems," *International Journal of Computer Applications*, vol. 68, pp. 12-18, 2019.

7. A. Patel and R. Sharma, "Context-aware voice assistants for smart homes," *Proceedings of the International Conference on Smart Devices*, pp. 45-50, 2020.

8. K. R. Choudhury and P. S. P. Bhat, "AI and machine learning for voice interaction systems," *IEEE Software*, vol. 34, no. 2, pp. 88-95, 2020.

9. Y. Guo, Z. Liu, and Y. Zhang, "Natural language processing techniques in AI-driven chatbots," *International Journal of Artificial Intelligence*, vol. 25, no. 5, pp. 788-799, 2021.

10. D. D. Nguyen, "AI-based speech interfaces: Design and applications," *Journal of Intelligent Systems*, vol. 32, pp. 112-118, 2019.

11. P. A. James, "Real-time response generation in AI chatbots," *Artificial Intelligence Research Journal*, vol. 40, pp. 254-262, 2021.

12. H. L. Singh and J. K. Agarwal, "Natural language understanding in AI-driven assistants," *Journal of Language and Computing*, vol. 36, pp. 191-205, 2020.

13. L. M. Trujillo and M. A. Garcia, "Personalization and context in AI chatbots for customer service," *Journal of AI and Robotics*, vol. 28, pp. 105-115, 2021.

14. W. T. Kuo, T. S. Chiang, and C. H. Wu, "Multimodal AI chatbots: Combining voice and image understanding," *Multimedia Tools and Applications*, vol. 80, pp. 2893-2905, 2021.

15. A. V. Verma, "Text-to-speech synthesis: Advances and challenges," *IEEE Transactions on Speech and Audio Processing*, vol. 10, pp. 556-567, 2020.

16. D. Y. Park, "Multimodal AI chatbots: Integration of voice and visual inputs," *Proceedings of the IEEE Conference on Human-Computer Interaction*, vol. 26, pp. 124-133, 2020.

17. J. P. Saravanan and S. R. Kumar, "Evaluation of AI-based conversational agents for healthcare services," *Journal of Healthcare Informatics Research*, vol. 14, pp. 21-30, 2021.

18. P. A. Williams, "Challenges in natural language processing for voice-based AI systems," *International Journal of Computational Linguistics*, vol. 23, pp. 17-27, 2020.

19. K. I. Yamanaka and T. T. Tomoda, "Voice synthesis systems for mobile AI assistants," *Mobile Computing and Communication Review*, vol. 24, pp. 20-29, 2019.

20. F. L. Nogueira and M. F. Lopes, "Improved speech recognition for multilingual AI assistants," *IEEE Transactions on Language Technology*, vol. 18, pp. 72-82, 2020.

21. T. G. Martin, "Voice-powered AI assistants in education: An emerging trend," *Journal of Educational Technology*, vol. 25, pp. 42-49, 2021.

22.A. H. Amiri and J. M. Wright, "Optimizing user experience in voice assistants using machine learning," *Proceedings of the International Symposium on Intelligent Systems*, vol. 32, pp. 48-55, 2021.

23.P. K. Lee, "A review on security challenges in voice-enabled AI systems," *IEEE Access*, vol. 9, pp. 2657-2665, 2020.

24. A. C. Moore, "Context-based response generation in voice AI systems," *IEEE Transactions on AI and Robotics*, vol. 35, pp. 120-128, 2020.

25.S. R. Mishra and A. P. Kumar, "Adaptive learning in voice assistants: A survey," *Journal of AI Research*, vol. 42, pp. 10-18, 2021.