An AI Powered Voice Assistant for Enhanced User Interaction (Voice-Bot)

Arul Natarajan[1]
(Assistant Professor),
Computer Science and Engineering
St. Peter's Engineering College
Hyderabad, India

Badam Shashank[2]
Computer Science and Engineering
St. Peter's Engineering College
Hyderabad, India

Vidya Rajasekaran[3]
Department of Information Technology
B.S. Abdur Rahman Crescent Institute of Science and
Technology, Chennai, India

Boga Roja[4]
Computer Science and Engineering
St. Peter's Engineering College
Hyderabad, India

Amjan Shiek[5]
Professor and Head, Computer Science and Engineering
St. Peter's Engineering College
Hyderabad, India

J. Ramesh Babu[6]
Associate Professor, Computer Science and Engineering
St. Peter's Engineering College
Hyderabad, India

Abstract:- AI Voice Assistant can undoubtedly make it easier for people who are fewer techs knowledgeable to use their computers and do activities. It lowers the barrier for folks who might feel frightened or unfamiliar with conventional computer interfaces by offering voice-based interaction. The user experience is made more convenient and accessible by the ability to launch programs using voice commands, such as "open Notepad" or "open File Explorer". It makes it easier to start particular applications without having to manually browse through menus or desktop shortcuts. Your AI voice assistant appears to offer a number of sophisticated functionalities in addition to standard computer functions like resuming and shutting down. The book recommendation can propose books depending on the user's choices, while the movie recommender can help consumers discover new films based on their preferences.

Keywords:- Deep Face, Voice Bot, Speech-to-text, and Guidance

I. INTRODUCTION

A desktop application is AI Voice Assistant. All computing apps and the vast majority of different works will function effectively in our daily lives. Natural Language Processing (NLP) is an area of AI that relates to software. Voice assistants can interact with users in natural language and utilize cloud computing. Using AI components like speech recognition, Python, text to speech, speech Engine, and Deep Face, it is a very helpful tool for completing voice tasks. It primarily consists of a section with recommendations for films and literature, the final module is a sophisticated one with an emotion-based music player that recognizes the user's facial expression and plays music in line with that emotion. You can communicate with your computer by using a voice assistant. It is software for Windows that includes an intelligent personal assistant, a human language interface, automation, and speech recognition.

When IBM Shoebox was introduced in 1962, they were able to launch their long history of voice assistants at the Seattle World's Fair. The tool can be used as a basic calculator and can recognize digits. Natural discourse Processing (NLP) methods are used in one application area of AI voice assistants to comprehend and interpret human discourse. They use cloud computing to process and analyze voice data, enabling natural language interaction with users. In general, AI voice assistants make use of NLP techniques, cloud computing, and have developed into useful tools for carrying out actions via voice commands, making them effective and user-friendly in a variety of sectors.

II. LITERATURE SURVEY

According to [1], voice assistants are a crucial component of our technological ecology since they allow us to communicate with gadgets and carry out activities orally. Nowadays, we all choose automatic system works over routine manual working processes because we require everything to be completed easily and faster than predicted. We become more efficient and get more work done thanks to automatic processes. In order to accomplish all of our tasks automatically, we have built this assistant in that manner. In [2] a practical and effective method for managing chores, voice AI has become a crucial component of our technological landscape. Users can carry out a variety of tasks utilising voice commands without the usage of typical manual workflows or manual input. This automation improves the user experience overall, promotes productivity, and saves time. It makes sense to extend the same procedure to include book recommendations. Similar collaborative

filtering techniques can be used to suggest books based on user choices and shared reading interests [3]. The work in [4] produces a customized musical expression that matches the user's emotional state by fusing facial expression analysis with music suggestion. The whole user experience is improved by adding a layer of interactivity and adaptability to the music listening process. This research [5] provided us with an understanding of the process used to play music based on a user's mood by observing their facial expression utilizing AI and DL algorithms. In addition to the aforementioned publications, we also found examples of desktop applications written in Python. Python's Tkinter library is used to create GUIs (Graphical User Interfaces). Python makes it a popular option for developers looking to quickly produce GUI-based applications.

III. PROPOSED SYSTEM

Given the numerous drawbacks of the current system, the suggested system is an AI-based application in which all computer applications operate with the aid of voice commands by using the user's speech and then react in accordance with the user's order. As a result, it saves time and energy. Additionally, instead of manually typing out numerous papers, a user can utilize voice commands to translate their words into text. They can use the voice command open [filename] to discover any files, directories, or application. When the user uses voice commands to get information, the voice assistant describes the data and only displays the relevant information. The ability to process mail using voice commands, log in using the user's voice, enter the data, and send the mail is another special feature. As a result, the voice assistant can be used for all other programs, including storing files and starting them. If the user is prepared to put in the effort, it can also suggest books to read and films to watch in addition to the advanced modules. The user encounters many circumstances throughout their daily lives. He or she may feel joyful, sad, afraid, surprised, or neutral in those circumstances. Deep face is a module that can identify a user's emotions and play music based on those emotions. The creation of a talk technique for the output is the final and most crucial module. This module includes a speak method that speaks out the output data for user commands and system-processed queries as well as for displaying the output. n Fig [1], we employ the Python Text to Speech (pyttsx3) module and the Microsoft Speech Application Programming Interface (sapi5) to turn the text into speech on the computer.
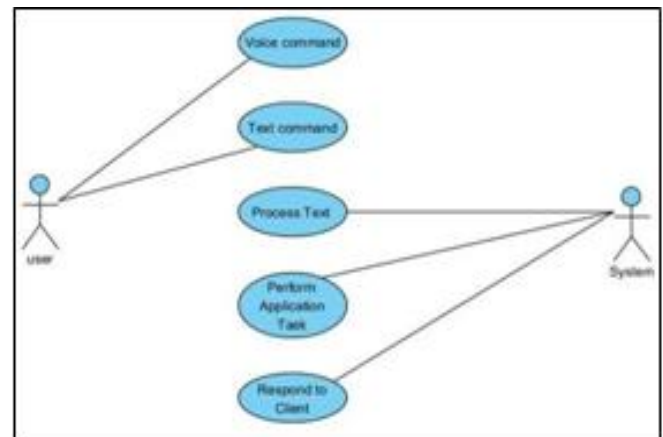

Fig 1 Use Case Diagram

When a user speaks a command into the voice bot, it runs the query and outputs the result. The function that voices assistants play nowadays is becoming important.

The ability of the final system to handle any kind of inquiries will help save a lot of time and work. It runs several applications, sends emails, and shows the user the date and time with the help of a few modules. The voice assistant is currently gaining popularity; in a few years, everyone will be using it, and its rapid expansion will continue.

IV. RESULTS AND DISCUSSION

There are various modules in the application. The "speech recognition" package that Python offers is a useful tool for implementing this feature. For audio recording, this library depends on another library. You can use the library's 'Recognizer' class to get started with voice command recognition. This course offers techniques for identifying speech from a microphone as well as other sources. The 'listen' technique is used to record audio input from the user through the microphone. To ensure reliable voice recognition, you might need to modify the recording's energy threshold to block out background noise. The ideal threshold can change based on the ambient noise level. Cloud-based systems like Google Speech Recognition, which need a live internet connection, are frequently used to translate speech to text. These services process and convert speech into text using powerful computers and machine learning algorithms. The module being discussed is a query processing system's implementation process. It accepts user commands, processes them, and generates the necessary outputs based on various conditions. The operations you mentioned are listed below: Speech Detection This module is in charge of turning oral commands from the user into text that the system can comprehend. It allows the computer to hear the user's voice and recognizes the command.

Condition Checking: To identify the best course of action, the module evaluates the condition depending on the received instruction. The system moves forward to produce the right output if the condition evaluates to true. Date and Time: To access the current date and time, the system uses the date-time module. As a result, the system is able to give

the user the current date and time upon request. In order to launch or open any application, such as Google, YouTube, Notepad, etc., we use the operating system (OS) module. The startfile() library and listdir() libraries are used to locate the location of the necessary file and to launch any application, respectively.

In Fig 2, the speech recognition module records the user's spoken instruction to "Open Notepad" and turns it into text. After conversion, the text is analyzed to see if the "Open Notepad" condition is true.
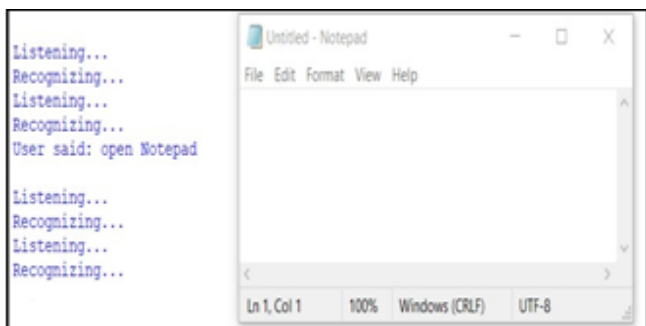
Fig 2 Notepad Page

The system then generates the corresponding result, which is to launch the Notepad application, because the condition is true. As a result, the user can interact with the Notepad application through the user interface.
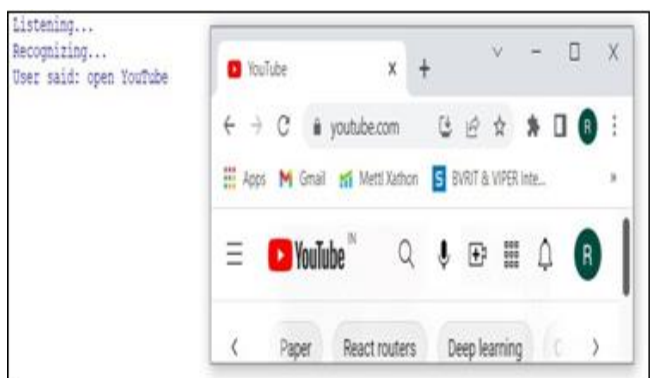

Fig 3 YouTube Page

Fig. 3 is the speech recognition module in this instance records the user's spoken command to "Open YouTube" and converts it to text. Following conversion, the text is utilized to test the condition, in this case "Open YouTube." The system generates the corresponding output after the condition is satisfied, which causes the YouTube website or application to launch.

As shown in Fig. 4, integrating the Deep face module, the system can determine the user's emotions from their facial expressions and make tailored music recommendations as a result.
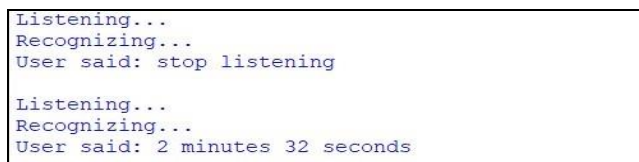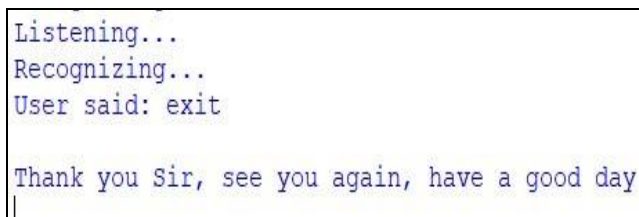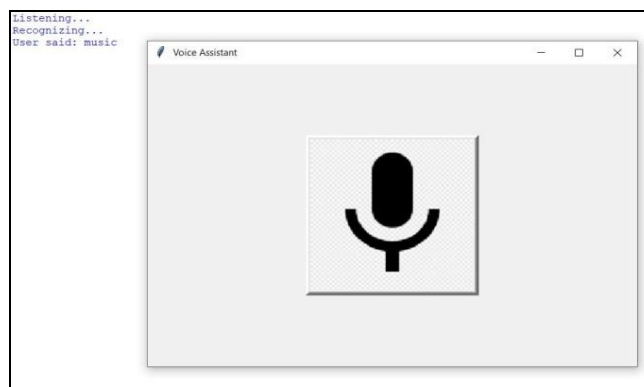

Fig 4 Execution Page

Fig 5 Response Page

Fig 6 Microphone Access

This feature gives the app a thoughtful and interesting touch while providing comfort and solace in trying circumstances.
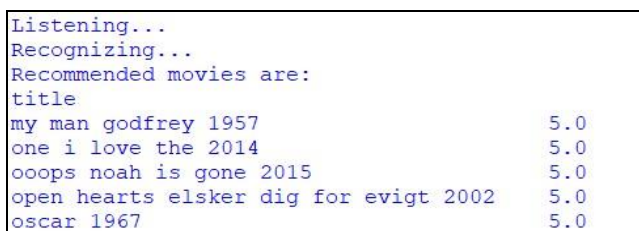

Fig 7 Movie Recommendation

Fig 7 provides accurate movie recommendations to the user by integrating a thorough recommendation algorithm that assesses many aspects. To produce initial recommendations, it considers elements including movie reviews, release dates, and viewership. The module also adjusts as the user watches more and more content, further honingthe recommendations to match their interests.

V. CONCLUSION AND FUTURE WORK

Voice assistants are likely to be used more frequently as they develop and become more well-liked. Voice assistants have the ability to completely change how people engage with technology due to improvements in natural language understanding and rising user acceptance. These assistants are important in daily life because they can

expedite a variety of tasks, from sending messages to buying tickets, and so save time.

However, issues including user preferences, technological developments, privacy concerns, and the creation of powerful and dependable voice recognition and processing skills will all have an impact on the adoption and use of voice assistants. Nevertheless, the rise in popularity of voice assistants portends a day when their use will be pervasive and provide users with major time-saving advantages. The primary goal of the project was to create a Desktop Assistant that would be used to find solutions to questions supplied by users. to give the user the information they need in a sufficient amount. Background investigation was done, including a rundown of the discussion process and any available relevant desktop Assistants.

Natural words Processing is the method used by voice-controlled devices to understand the human speaker's words, process the question, and answer to the human with the outcome. Understanding the gadget means that artificial intelligence must be incorporated into it in order for it to function intelligently, operate IoT applications and devices, and reply to queries that search the web for answers and process them. Even though they are bold ones, technology is currently taking its first steps. However, we anticipate that personal assistants will advance quickly, providing better interactivity, better speech recognition, and the capacity to handle more complex issues.

Personal assistants will keep getting better. With each request that is completed, the system gets better thanks to machine learning algorithms. As a result, the speakers will actually advance in intelligence. This implies that they would be able to carry on conversations in addition to simply responding to individual queries.

Additionally, the intelligent voice assistant will pick up on the owner's preferences and routines. It will therefore be able to give a more tailored experience by anticipating inquiries and providing more accurate search results. The biggest barrier preventing the global adoption of smart speakers is localization. The majority are presently offered in Chinese, German, or English. The gateway to larger markets will be opened by localization to other languages.

The capabilities of voice assistants are constantly evolving. To enhance the capabilities of assistants, AI businesses leverage data from current systems. In the end, according to Lucas, the voice assistant may become so intelligent that it will automatically order a pizza if you mention being hungry. It will determine that stating that you are hungry is equivalent to placing an order for pizza using historical data from your prior transactions.

REFERENCES

[1] Hoy, M. B. (2018). Alexa, Siri, Cortana, and more: An introduction to voice assistants. *Medical Reference Services Quarterly*, *37*(1), 81-88. https://doi.org/10.1080/02763869.2018.1404391

[2] Majumdar, S., Kirkley, S., & Srivastava, M. (2022). Voice command AI assistant for public safety. *2022 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies(3ICT)*. https://doi.org/10.1109/3ict56508.2022.9990716

[3] O'Shaughnessy, D. (2003). Interacting with computers by voice: Automatic speech recognition and synthesis. *Proceedings of the IEEE*, *91*(9), 1272-1305. https://doi.org/10.1109/jproc.2003.817117

[4] M. Chenna Keshava, P. Narendra Reddy, S. Srinivasulu, & B. Dinesh Naik. (2020). Machine learning model for movie recommendation system. *International Journal of Engineering Researchand*, *V9*(04). https://doi.org/10.17577/ijertv9 is 040741.

[5] Emotion recognition from facial expression using deep learning. (2019). *International Journal of Engineering and Advanced Technology*, 8(6S), 91-95. https://doi.org/10.35940/ijeat.f1019.0886s19

[6] Kenny, P. G., & Parsons, T. D. (2011). Embodied conversational virtual patients. *Conversational Agents and Natural Language Interaction*, 254-281. https://doi.org/10.4018/978-1-60960-6176.ch01 1

[7] Epstein, J., & Klinkenberg, W. (2001). From Eliza to internet: A brief history of computerized assessment. *Computers in Human Behavior*, *17*(3), 295-314. https://doi.org/10.1016/s0747-5632(01)000 04-8

[8] Kunekar, P., Deshmukh, A., Gajalwad, S., Bichare, A., Gunjal, K., & Hingade, S. (2023). AI-based desktop voice assistant. *2023 5th Biennial International Conference on Nascent Technologies in Engineering (ICNTE)*. https://doi.org/10.1109/icnte 56631.2023.10146699

[9] Jacobsen, N. (2016). IPhone 4S: Ein gigantischer, Aber trügerischer Erfolg. *Das Apple-Imperium 2.0*, 51-54. https://doi.org/10.1007/978-3-658-09548-2_11

[10] Han, Y. (2021). A study on communication between AI (Artificial intelligence) voice assistant and AI speaker users and services usage : Focusing on anthropomorphism personality, social presence, and personalization. *Journal of Communication Science*, *21*(3),225-275. https://doi.org/10.14696/jcs.2021.09.21.3.225

[11] Naureen, A., Siddiqa, A., & Devi, P. J. (2022). Amazon product Alexa's sentiment analysis using machine learning algorithms. *Lecture Notes in Networks and Systems*, 543-551. https://doi.org/10.1007/978-981-16-8512-5_57