IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2347-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 4, April 2024

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.423

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304201 |

Virtual Voice Assistant Using Large Language Models - Ultron

A. Sai Prashanth, Dr. S. Siva Skandha, B. Sivaiah

UG Student, Department of CSE, CMR College of Engineering & Technology, Hyderabad, Telangana, India

Associate Professor, Head of the Department of CSE, CMR College of Engineering & Technology, Hyderabad, India

Associate Professor, Department of CSE, CMR College of Engineering & Technology, Hyderabad, India

ABSTRACT: The voice assistants have revolutionized the human-technology and information world. They can be found in mobile phones, artificial intelligence-based speakers or even some vehicles, all seamlessly blending with daily lives. Siri, Google Assistant and Amazon Alexa are examples of these virtual helpers performing additional functions such as information retrieval, reminders and device control through natural language. Moreover, they help disabled people gain access to their iPhones and enjoy life more. In its evolution towards more intuitive conversation based interfaces, voice assistants will reshape interactions between humans and devices. "Ultron – Voice Assistant" is a web application that allows users to interact with a chatbot by using voice commands or text input. The Web Speech API is responsible for converting spoken words into text while the backend uses Django to handle user queries. It incorporates PaLM model which generates intelligent replies. The chatbot's feedback can be seen on interface and also played back using Google Text-to-Speech option. This assistant can identify who the user is and can talk to him basing on his mood as well as situation he's in at the moment. Therefore it has) a greater impact on (security) user experience than any other factors do . Additionally, this assistant offers services like capturing screenshots, cracking jokes or launching different.

KEYWORDS: Artificial Intelligence, Large Language Model, Voice Assistant, Python, Chatbot, Real-Time Audio Descriptions, Virtual Assistance, PaLM.

I. INTRODUCTION

In this ever-evolving landscape of technological advancements the demand for innovative human-computer interactions is a great paramount. The traditional interfaces often disrupt and even slow. The dynamic fast-growing environment requires the dynamic virtual assistants, which can perform the tasks at an ease. The modern Virtual assistants are limited to certain feature and the complete virtual feel is missing, to improvise this the advanced innovative solution is needed. By integrating the cutting-edge speech recognition technologies, the application will interpret user inputs accurately and respond dynamically with informative responses. The integration of the pre-trained model PaLM provides the most accurate responses. This project's significance lies in its ability to bridge the gap between humans and machines, by offering an innovative and efficient solution to enhance productivity and accessibility.

II. RELATED WORK

The primary research objective of this project is to develop a Voice Assistant system that seamlessly integrates the Natural Language Processing, Machine Learning and web-based functionalities and provide the accurate and efficient responses. By harnessing the trending technologies such as Large Language models, face recognition, speech recognition and other APIs and real-time communication services like WhatsApp, the aim is to create a multifunctional assistant that can not only comprehend and respond but also perform various tasks. There are various existing solutions that are present.

A. **Google Assistant.** Google's Voice assistant, generally known as Google Assistant [1], it is an extensively integrated solution that operates across wide range of devices and platforms. It enables voice forward controls in devices, Using the assistant we can launch applications, perform various tasks and access content. It can be activated by using the command 'Hey Google', "Ok Google". The Google assistant enables users to launch apps and performs tasks such as launching features from the assistant, displaying app information, and also can suggest voice commands. The assistant integrates with Google's ecosystem of services [2], such as Google search, Google Maps, and Calendar. Although with lot of

| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304201 |

useful features it also lacks in an integrated screen. Emails and messages cannot be read or sent.[3-4] The Google assistant is poor in automation ad it cannot automate the mails and messages. It cannot provide the actual rather it just gives the related content websites. The Humanly chat feature is not present in it.

B. **Amazon Alexa.** Amazon's voice controlled assistant, Alexa, works on devices like Amazon Echo. Alexa's boasts capabilities such as playing music, setting reminders, providing weather updates, and also controlling smart home appliances. Alexa is cloud based voice assistant which is build on natural language processing and can do wide range of functionalities which includes smart home, shopping, entertainment, news, communications, and more [5]. Amazon alexa also offers the custom model to create own product using the custom models. The Amazon skill kits [6] is a software development which enables developers to create applications which are known as Skills. It is also one of the services of the Amazon Alexa. Even with this advancements and features it lacks in poor individual advise and low sound quality [7].

C. **Apple Siri.** Apple Siri is the voice and digital assistant that is the part of the Apple Inc [8]. It is available on iOS devices and macOS. Siri offers voice recognition feature and natural language understanding, allowing users to perform various tasks like sending messages, setting remainders, and opening applications. It uses voice queries and gesture based controls. It is developed by the SRI Internation Artificial Intelligence Center. The speech recognition engine was given by the Nuance Communications [9-10]. Siri's merit lies in its integration with Apple's ecosystem, and enhancing the user experience across Apple devices. The challenges include difficulty in analyzing complex queries and adapting to various accents. Siri will not function if the user is with poor internet connection [11].

D. **ChatGPT.** The ChatGPT [12-13] is a revolutionary chatbot which is developed by the OpenAI. It is based on the large language model and it enables the users to refine the conversation and can give humanly feeling while using. It is a Generative AI, which provides responses based on the queries accurately and efficiently. It has become the fastest growing consumer software in the world by gaining over 100 million users worldwide. Currently the Generative pre trained transformer model if fined tuned and versions of the gpt are being released [14]. With the large advantages and uses the ChatGPT has certain limitations such as it cannot control the devices and cannot have access to the system.

Comparison Dimension	Google Assistant	Siri	Alexa	ChatGPT
Voice Recognition	High	High	High	Moderate
NLU	Excellent	Good	Good	Moderate
Integration of Devices	Extensive	Limited	Extensive	Limited
Third-Party App Integration	Moderate	Limited	High	Limited
Conversation Depth Context	Limited	Limited	Limited	High
Opening apps	Yes	Yes	Yes	No
Music Playback	Yes	Yes	Yes	No

Figure 1. Comparison between various Assistants and Technologies.

III. METHODOLOGY

A. High-level methodology

The proposed solution is a Voice Assistant application using Django. The objective of the proposed solution is to develop a voice controlled assistant that can perform various tasks based on the voice as well as textual commands and perform operations such as giving generative responses, opening applications, providing information, navigation, sending messages, playing music and more. The assistant can understand and remember user-specific information, such as name, favorite songs, and more. The application can even allows the users to set the timers by specifying hours,

| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304201 |

minutes and seconds. And it has to notify the user. The application is responsible for providing audio as well as textual output. In order to achieve this, there are several technologies that have to be integrated and used, they are:

B. **PaLM gemini-pro-vison:** Gemini is the Google latest generative model of the PaLM family [15]. There are several gemini api which have their own functionalities, such as Gemini 1.5 Pro deals with both text and images data, Gemini 1.0 Pro deals with only textual data. We have used the Gemini 1.5 Pro which deals both text and images and created an api for that [16]. This model performs well at a variety of multimodal tasks such as visual understanding, classification, summarization, and also create content from images and videos. The main purpose of this model is for getting the accurate and fast responses, like in google assistant it only provides the website which are similar to the searched content this model gives generative answer that can be stored and based on that query we can continue the conversation.

C. **Python:** Python [17] is the open source high level language which can be used for various purposes. Python contains the large number of libraries and huge community which mode it efficient and powerful language. There are various Machine learning and Artificial Intelligence based packages which are responsible for model training and model deployment. Python contains the internal Speeach recognition package which is used for speech recognition and voice output. We used Python programming language due to its vast range of Libraries. We used several libraries in python such as: pywhatkit – it is used for opening the whatsapp and sending messages. This library is used for automation of whatsapp messages. AppOpener – this library is used for opening the applications that are present in the system. It integrates with the operating system and searches the specified application and opens them. pyautogui [18]- this library is used for moving mouse and clicking, sending keystrokes to the application, taking screenshots, locating the application window, and displaying the alert messages. Wikipedia- this library is used to extract the information that is present in the Wikipedia and send to the user.

D. **Django:** Django [19] is a high-level Python web framework that is used for rapid development and clean, pragmatic design. It is a free and open-source, and it is Python-based web framework that runs on a web server. It follows the model–template–views architectural pattern. It helps the developers to complete the task quick and more efficiently. It also provides the security and helps the developers to avoid common mistakes. We initialized the Django project using the terminal and created the virtual environment, and installed all the packages, then created the Django application. Then created the models, views and urls. The Django gave the python code a framework and cross platform support. The Html templets are rendered and created a user friendly interface. The Django also gave the auto authentications using the internal authentication support. Thus it is made easy for the development of the interface.

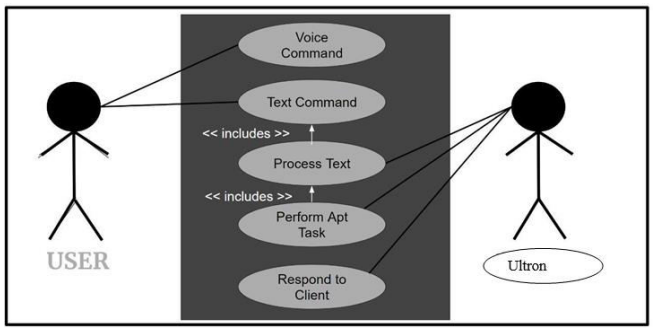

Figure 2. Representing the usecase diagram

E. **Text-to-Speech:** Text to speech is the important aspect in a voice assistant. The primary objective of the voice assistant is to speech with the system and the system speaks to us. In order to convert the text to speech we used the Google text to speech modules which is very efficient and can provide the output with general accent. While writing the code provided the option such that the user can choose whether to hear or just need the textual output, he can turn off the mice symbol to cancel the audio output.

IV. EXPERIMENTAL RESULTS

In this section, we present the outcomes of our project's implementation and engage in a discussion around its implications. By examining the results obtained by the solution the effectiveness and the potential of our system is assessed. Through the concise analysis, we explore collective impact on achieving the objectives and metrics. We further improve the areas where the performance is lagged and ultimately aiming to contribute the advancements in the

| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304201 |

generative models and voice assistants. The model which we created performed well and gave optimal output with audio outputs.

Performance metrics

Code Organization and Structure: The code which we have written demonstrates effective organization, clearly separating the static files such as HTML, Javascripts and python sections. The external libraries that are used and frameworks, such as Django, jQuery, and Font awesome, enhances the functionality and styling and enhanced the user experienced. The user interface and user experience is enhanced by the UI elements and it is the clear arrangement, contributing to user friendliness. This voice assistant interface includes the intuitive buttons for actions like text box for entry, speaker, microphone and toggle buttons and output box.

Voice Recognition and Response:

Approach: The project leverages the speech recognition library of python for voice recognition purposes. Speech is synthesized and the capabilities are employed through the python's pyttx3 library, and offering audio responses to the users.

Django integration: The Django integration makes the process easier and more effective. It is responsible for integration of external services, The services such as Wikipedia search, and summary retrieval have been integrated into the project. The pywhatkit library advanced the automation of messages.

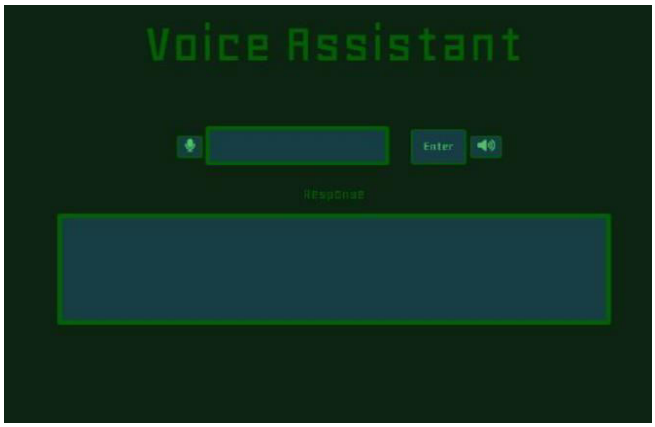

Figure 3. Representing the UI of the proposed system.

Chatbot integration: The Google's Gemini pro vision model enhances the user interactions through chatbot responses. The model is capable of turning the unstructured data into structured data, it can also adjust the prompt and provides the structured responses. The obtained responses are then stored as the json format. Then the output is displayed by using the representation state transfer. The Ajax is used to manage the asynchronous interactions between the frontend and the backend i.e; Django.

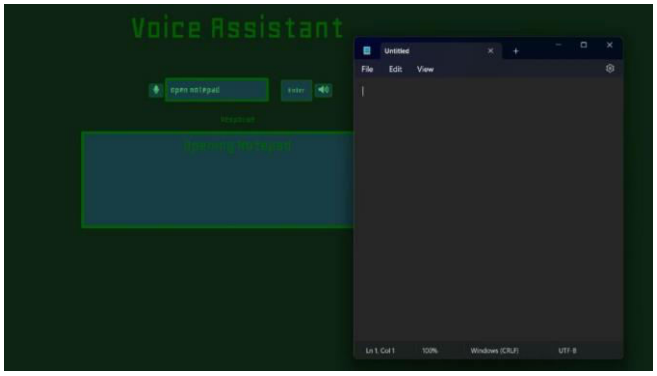

Figure 4. Opening application using Voice assistant.

| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304201 |

The error handling is done effectively such that the adequate verification of speech recognition and synthesis library availability prior to their use is demonstrated. The code is handled with possible errors and adequate for recognized and unrecognized speech. The user data handling includes like maintaining the user preferences like favorite songs, names are stored in the cache memory of the Django application and updated as needed. The effectively manages the user preferences and data persistence across interactions at various levels.

Comparison:
The existing solutions often struggle with various factors. Each existing system has its own flaws. The Google assistant cannot provide the generative and logical outputs. It is the major flaw in the Google assistant and it cannot understand the complex queries and maintain descriptive responses. Where as in Alexa, it also cannot provide the logical reasoning and generative responses. Contrary to these ChatGPT provides the excellent descriptive and logical reasoning, but it cannot open applications and do tasks. So overcome all these our system can effectively performs all types of tasks. It can open any application and does tasks and also provides the descriptive and logical reasoning responses. It can even takes screenshots, play music and can automate the messages which cannot be seen in any other existing solutions.

Integration:
To make the voice assistant project work, it has to synthesize different technologies and libraries into one interactive interface that enables user communication with a voice assistant via commands. HTML is used for the user interface while JavaScript serves the purpose of handling user interactions as well as speech recognition. On the other hand, Django is employed in the back-end processing and API interactions.

A Step-by-Step Process of the integration are:
1. HTML Interface (index.html):
This HTML interface has buttons for starting voice recognition, switching on/off text-to-speech function and sending user's input. The interface includes a textarea for bot responses. Javascript code is in an html file where it handles user interaction with django backend.
2. JavaScript Functionality:
When the `start listening` button is clicked by the user, it will invoke `startListening()` function.StartListening() initializes the recognition object from Web Speech API which listens to what users say. It then turns this recognized speech into text that populates into an input field. User can put by typing or using enter key after writing words in such a way that triggers event $"#rec".click().This entered text will be sent
3.Django Backend (Python):
In this case, Django backend is where the user input from AJAX POST request goes to. Through that input, it's able to generate a request with OpenAI Chat API for a bot response. The response of the bot is gotten from Gemini API. Such JSON object will be sent back to the front end as a response.

Overall Interaction: Accordingly, users use buttons on HTML interface which they click before supplying voice or text inputs. JavaScript handles voice recognition, user input and sends data at the backend by calling Django functions. The users' commands are processed by Django in order to talk to external APIs so that responses can be obtained as a result of such requests made by bot. Incase speaker is set on then frontend uses TTS synthesis for saying chatbot response out loud.

V. CONCLUSION

In this regard, it can be concluded that the presented project of Ultron - Voice Assistant demonstrates how modern technologies can be combined to create an efficient AI companion which is also user-friendly. The aim of this project is that by unifying voice recognition, text-to-speech, and language models from OpenAI, users will be able to have a normal conversation with their assistant. Some of the features included in this project include; emotion detection, application control and personalized interactions among others making it more intuitive and personal.

Additionally, through backend management using Django as well as employing Tkinter library in GUI design, the project becomes more complex. In other words, with capability such as searching webpages and Wikipedia queries or even sending messages then by all means you only require a flexible tool fit for numerous types of users. It indicates that building an AI-driven assistant itself already holds promise for future endeavors yet opening up new worlds in its ability to create more sophisticated digital voices that are taking into account ever-changing artificial intelligence

| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304201 |

patterns. In conclusion, when we evaluate possible future developments on this project some exciting opportunities emerge. Initially, enhancement of contextual sensitivity for the assistant may lead to natural conversations.
Another way to improve on these models could be by enabling them to understand and give responses to complex user queries, using context from ongoing conversations.

Furthermore, it would be valuable to expand the assistant's ability to comprehend and execute multi-step commands. That is where building more advanced dialogue management systems that handle task sequences, remember previous interactions and keep context over time during long conversations may come in handy. In addition, incorporating external APIs and services could enable tasks like booking appointments, ordering food or controlling smart devices hence making the assistant a true virtual concierge.

Besides that, one will find out that this new generation of assistants can answer back with empathy by combining them with voice recognition technology which can recognize human emotions. This will also involve adding some recommendation system for suggesting personalized music videos or other content based on user preferences and past interactions.

To sum it all up, what lies ahead seems to be a future that has the potential of refining as well as expanding upon this Voice Assistant project such that it becomes an advanced AI companion that is intuitive enough to fit in effortlessly into people's lives.

REFERENCES

[1] https://developer.android.com/guide/app-actions/overview.
[2]https://developer.android.com/guide/topics/connectivity/cross-device-sdk/overview .
[3] Comparative Analysis of Voice Powered Assistants by Dr. Sangita Jaybhaye1, Akshay Manchekar2 , Amogh Dixit3 Ayush Ingle4 , Fatima Khatib5 1, 2, 3, 4, 5Department of Computer Engineering, Vishwakarma Institute of Technology, Pune, India
[4] https://www.mercurynews.com/2013/03/07/strengths-and-weaknesses-of-apples-siri-and-google-now/
[5]https://www.amazon.com/b?node=21576558011&ref_=alxcom_lrnmore_btn_23
[6]https://developer.amazon.com/en-US/docs/alexa/ask-overviews/what-is-the-alexa-skills-kit.html
[7] https://downdetector.in/status/amazon-alexa/
[8] "Use Siri on all your Apple devices". support.apple.com. November 2023.
[9] "Report of unscheduled material events or corporate event » Microsoft 8K". April 12, 2021. Retrieved April 14, 2021.
[10] "SCANSOFT, INC. (nuance Communications, Inc.)". www.sec.gov. Retrieved April 12, 2021.
[11] https://botpenguin.com/blogs/pros-cons-and-use-cases-of-using-siri-on-your-apple-device
[12] "ChatGPT – Release Notes". Archived from the original on January 12, 2024. Retrieved January 16, 2024.
[13]https://help.openai.com/en/articles/6825453-chatgpt-release-notes
[14] Gertner, Jon (July 18, 2023). "Wikipedia's Moment of Truth". *The New York Times Magazine*. Archived from the original on July 20, 2023. Retrieved July 19, 2023.
[15] https://ai.google.dev/models
[16] https://ai.google.dev/api/python/google/generativeai
[17] https://opensource.com/resources/python
[18] https://pyautogui.readthedocs.io/en/latest/
[19]"django/README". *GitHub*.Retrieved 8 September 2020.
[20] https://www.dataquest.io/blog/python-api-tutorial/
[21] https://cocodataset.org/#download

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details