

# AI Voice Assistant

Onkar Yede<sup>1</sup>, Srushti Raut<sup>2</sup>

<sup>1</sup>Student at MAEER'S MIT College of Railway Engineering and Research, Barshi, Maharashtra, India.

<sup>2</sup>Assistant Professor, Dept. of Computer Science and Engineering, MAEER'S MIT College of Railway Engineering and Research, Barshi, Maharashtra, India.

\*\*\*

**Abstract** – In the imminent future, all electronic devices are anticipated to function with the aid of a remote assistant, which is easily accessible yet requires a certain level of vulnerability. This configuration permits users to interact with the system via voice commands. Users can instruct the assistant to perform various tasks within the framework's capabilities, such as playing music, opening applications, accessing tabs, visiting websites, and more. Expert programmers, known as voice associates, possess the ability to comprehend human speech and respond with carefully crafted responses. The primary objective of our voice assistant is to educate individuals and provide prompt, precise solutions. Voice assistance receives input from microphones (both wired and Bluetooth) and translates it into comprehensible English for the computer system, subsequently delivering the necessary responses and solutions to the user. The assistant establishes a connection to the Internet to fulfill user queries, facilitated by a natural language processing algorithm that enables communication using various forms of human language.

**KeyWords:** TTS (Text-to-Speech), AI (Artificial Intelligence), Voice Assistant, NLTK(Natural Language Toolkit),Python.

## 1. INTRODUCTION

Some of the most recognizable voice assistants include Siri by Apple Inc., Alexa by Amazon, Cortana by Microsoft, and others. These assistants are software agents designed to understand and respond to spoken language with synthesized voices. The transition from the futuristic concept of talking to computers to its realization is evident in today's technological landscape. Advancements continue to occur rapidly, driven by the persistent efforts of computer scientists. Since the inception of computers, there has been a desire to communicate with them, a desire that has now been fulfilled through the technologies developed by dedicated researchers. In recent years, numerous products have been introduced, making voice assistants more accessible and integrated into our daily lives. These advancements have made voice assistants more affordable and user-friendly. As time progresses, additional features and platforms are continuously being added and updated, enhancing the utility and versatility of voice assistants. In today's rapidly evolving society, there's a widespread preference for efficiency and speed in

completing tasks. To meet this demand, we're embarking on the development of an AI Voice Assistant aimed at streamlining various activities. Whether it's opening files or conducting Google searches, this assistant will cater to every need, promising convenience and promptness in task execution.

Furthermore, this personalized assistant will imbue users with a sense of authority and control, akin to having a dedicated aide at their disposal. By utilizing voice commands, users can effortlessly convey instructions, which the assistant will interpret and act upon accordingly. This seamless interaction enables users to engage with the machine in a natural, conversational manner, enhancing user experience. The primary objective of this endeavor is to harness the power of Natural Language Processing (NLP) to create an AI Voice Assistant that serves as a valuable tool, particularly for individuals with visual impairments or physical disabilities. By providing a hands-free solution, this assistant aims to simplify daily tasks and enhance accessibility, thereby empowering users to navigate their routines more effectively and independently.

### 1.1 Project Scope

The pace of change in technology has been remarkable. Reflecting on the past two decades, it's clear how rapidly Voice Recognition has evolved. In its nascent stages, it was merely a concept, a distant dream of achieving full-fledged interaction with computer systems. Today, however, that dream has materialized into reality. We can now converse, inquire, and even delegate tasks to our devices.

This technological advancement has propelled the world forward, revolutionizing the way we interact with machines. While current virtual assistants are already accessible and beneficial, there remains ample room for improvement. Their comprehension and reliability, especially in critical situations, still require significant enhancement.

Looking ahead, the virtual assistants of the future will be equipped with cutting-edge artificial intelligence technologies, including machine learning, neural networks, and the Internet of Things. These advancements promise to elevate our capabilities to unprecedented levels. The

potential of virtual assistants extends far beyond our current achievements, offering opportunities to reach new heights and accomplish tasks previously unimaginable.

## 1.2 Project Goals and Objectives

### 1.2.1. Project Goals:

The main objective of developing an AI Voice Assistant Project for desktops is to offer users a hands-free and convenient tool to accomplish various tasks on their computers without manual input. Leveraging Natural Language Processing (NLP) algorithms, this assistant can comprehend user voice commands, interpret them, and carry out actions accordingly. Integration with diverse software and applications installed on the user's computer enables tasks like opening applications, web searches, accessing YouTube, and utilizing text-to-speech features, all via voice commands.

This functionality proves especially advantageous for individuals with mobility impairments, for whom traditional keyboard and mouse interactions may pose challenges. Ultimately, the aim of creating an AI Voice Assistant Project for desktops is to simplify users' daily routines, boost productivity, and elevate their overall computing experience.

### 1.2.2. Objectives:

- i. The objective is to design a system optimized for desktop usage, prioritizing efficiency in operation.
- ii. The aim is to develop a user-friendly software tool that is both accessible and easy to use.
- iii. The goal is to create an interactive software solution that is cost-effective.
- iv. Ensuring that the AI Voice Assistant accurately recognizes and responds to a broad spectrum of voice commands, accommodating diverse accents and speech patterns, is essential for enhancing user experience and accessibility.
- v. Integration of the AI Voice Assistant with commonly utilized desktop applications like email clients, web browsers, and productivity software is crucial for providing a seamless user experience.
- vi. Conducting user testing and soliciting feedback from visually impaired and handicapped users is imperative to guarantee that the AI Voice Assistant is accessible, user-friendly, and tailored to their specific requirements. This process will validate the effectiveness and utility of the AI Voice Assistant for this target demographic.

## 2. LITERATURE REVIEW

Artificial intelligence (AI) systems capable of facilitating natural human-machine interaction, encompassing speech, gestures, and facial expressions, are experiencing a surge in popularity. The evolution of interaction dynamics, where machines interpret real human language, has been a focal point of extensive research and widespread interest. Today, machines are not just learning to understand human language; they are observing human activities, routines, and behaviors to become personalized assistants, shifting the dynamic from machine-serving-human to the machine becoming the user's assistant.

Personal assistants have become integral to our daily lives, often without us fully recognizing their significance. This is primarily due to their multifaceted features and user-friendly interfaces, which significantly enhance ease of use and functionality. A user can streamline their focus on priorities by leveraging a personal assistant to automate routine tasks. These assistants offer a range of features such as making calls, sending messages, capturing photos, managing to-dos on the fly, and browsing the internet. By utilizing these features, individuals can save considerable time and effort, allowing them to allocate more attention to what truly matters, whether in their personal or professional lives. Personal assistants excel at automating mundane tasks that often consume significant time.

In unfamiliar work environments, individuals frequently struggle to locate necessary applications, such as browsers or IDEs, leading to wasted time in searching. A voice-activated personal assistant can alleviate this challenge by automating the process. With a simple voice command, users can prompt the assistant to locate and launch the desired application, eliminating unnecessary time wastage.

This paper explores the benefits of utilizing a voice-activated personal assistant to expedite tasks and enhance user convenience. By responding to voice instructions, these devices facilitate quicker and easier task execution, ultimately improving overall productivity and efficiency. The final product of an intelligent assistant is shaped by the distinctive methods and approaches employed by each organization utilizing it. While one assistant may excel in executing tasks accurately and in line with user preferences, another may prioritize precision and efficiency, requiring minimal explanations or corrections. It's important to acknowledge that no single assistant can perform flawlessly in every aspect.

The traits and capabilities of an assistant are primarily determined by the focus areas emphasized by its developers. The data sources utilized, whether search engines, diverse information repositories, or social networks, exert a significant influence on the assistant's capabilities. The breadth and quality of data from various

sources contribute to the effectiveness of the assistant's outcomes.

While there exist diverse learning strategies, algorithms, and methodologies, the fundamental principles underlying the development of such systems remain largely consistent across the board. Ultimately, the success of an intelligent assistant hinges on the integration of relevant data sources, robust algorithms, and effective learning mechanisms to deliver optimal performance. The initial approach we examined involves acoustic analysis, which entails capturing and analyzing partially pronounced features using learned data to complete them. The user's command is segmented into individual words, cross-referenced with a command list, and then a corresponding response is generated based on the identified command. [1]

The second assessment pertains to a smartphone application featuring several functionalities, which integrates natural language processing for audio-to-text conversion and employs a decision tree algorithm to classify photographs and identify objects within them. The application operates by extracting keywords provided by the user, which are then matched against training data to generate a response. [2]

The models under examination encompassed various task or service categories, each associated with a class of service modules. These modules offered identical services and presented consistent interfaces, yet exhibited differences in performance, computational complexity, memory usage, portability (with some relying on proprietary services), and potential dependencies on external cloud services. [3]

The assistant being evaluated is resource-efficient, interactive, and adaptable. The developed prototype operates effectively on an affordable Raspberry Pi 3 device. For testing, the system was integrated into an open-source home automation environment and continuously operated for several days, encouraging user engagement. The results demonstrated its accuracy, reliability, and user appeal. [4] Dialogue systems that integrate multiple user input methods, such as speech, image, video, touch, and manual gestures, are referred to as multi-modal dialogue systems. The foundational framework of generic dialogue systems encompasses various elements including the ASR Model (Automatic Speech Recognition), Gesture Model, Graph Model, Interaction Model, User Model, Input Model, Output Model, Inference Engine, Cloud Servers, and Knowledge Base for response generation. [5]

The examined DIVA assistant design operates as a strategy incorporating accent recognition, speech registration, and affording users increased manual control within the framework of a mobile application. To fulfill user objectives effectively, the virtual assistant employs

techniques such as natural language processing, speech pattern recognition, and machine learning. [6]

The subsequent system utilizes a meticulously curated dataset, which has undergone thorough cleaning processes. This cleaning involves the removal of punctuation, extraneous symbols, and other unnecessary elements. Furthermore, redundant logos were identified and eliminated from the dataset using the Bag of Words technique, following comprehensive data preparation. Despite diligent efforts to refine incoming data and transform it into coherent phrases and clear terminology, duplication has been diligently removed from our dataset. [7]

The chatbots examined in our study utilize natural language processing (NLP) techniques to understand user commands and provide responses. Their primary architecture comprises deep neural networks and APIs for text-to-speech and speech-to-text conversion. Commands in the form of voice signals, converted to text, are transmitted over a Bluetooth network to the robot, which employs OpenCV for processing. Additionally, the chatbot utilizes stemming, lemmatization, and word filtering in tandem with the NLTK module to conduct natural language processing and generate appropriate responses to user queries. [8]

The NLP-based chatbot under examination is designed to accommodate audio notes from users, in contrast to traditional text-based input methods used by other chatbots. This architecture initially converts the audio notes into text format. The main structure of the chatbot comprises deep neural networks and APIs for both text-to-speech and speech-to-text conversion functionalities. [9]

The most recent system being studied is a mobile application that depends on voice commands provided by the user. These commands are processed by the mobile device using natural language processing (NLP) techniques. Serving as a central hub, the mobile device decides which appliance should carry out specific actions to meet the user's request. Data transmission is handled through cloud infrastructure, which processes the information. By leveraging an Arduino Board, the mobile device establishes connections with appliances, enabling the Internet of Things (IoT) framework. [10]

### 3. PROBLEM STATEMENT

We're embarking on an AI-based Voice Assistant implementation utilizing NLP techniques. Our primary objective is to assist visually impaired or handicapped individuals by offering a hands-free tool, enabling them to access technology without requiring manual input. While existing solutions like Google Voice Assistant, Siri, and Cortana are widely known, they are limited to specific devices. Our aim is to develop a voice assistant that can seamlessly operate across all platforms, ensuring universal



accessibility for users regardless of their device preferences.

#### 4. PROPOSED SYSTEM

Our goal is to develop software that is both simple and highly effective. The proposed software is primarily designed to facilitate hands-free operation and streamline the user's daily tasks. It aims to save users time and money by providing efficient assistance. This desktop-based voice assistant operates by interpreting voice commands from the user and generating the desired output accordingly.

The assistant comprises three key modules. First, it accepts voice input from the user or client. Next, it analyzes and maps the user's input to determine its intent and function. Finally, the assistant verbally communicates the outcome to the user. The assistant will commence by receiving user input in the form of voice commands. Upon receipt, it will translate the analog voice input into digital text. If the translation is unsuccessful, the assistant will prompt the user for input once more. Once successfully converted, the input will be analyzed and mapped to a specific function.

Utilizing NLP algorithms, the assistant will convert the user input into machine-understandable language and associate it with appropriate responses. Subsequently, the output will be delivered to the user via voice commands. To facilitate this process, the assistant employs Text-to-Speech (TTS) modules to convert voice commands into text and text responses back into voice commands.

Additionally, users have the flexibility to customize the assistant's name according to their preference. This proposed assistant serves as a valuable tool for simplifying the user's day-to-day desktop tasks. The assistant swiftly searches and opens any designated folder upon receiving a voice instruction, aiding users in locating misplaced folders efficiently. It seamlessly accesses Google, Wikipedia, YouTube, plays music, and plays videos within the system in response to voice commands. Users can inquire about and adjust the temperature, date, and time simply by asking the assistant.

Moreover, the software can read aloud the text of a document upon request. When users ask for news updates, the assistant not only reads the news aloud but also displays it line by line, allowing users to control the reading speed. Additionally, the assistant offers Speech-to-Text conversion, enabling users to document text without the need for manual typing. This feature proves invaluable for tasks like composing emails and texts, as users can dictate their messages without using the keyboard.

Furthermore, users have the option to customize the voice of the assistant, selecting either a male or female voice based on their preference and comfort. Additionally, whenever the user summons the assistant, it greets the user by their username and adjusts the greeting according to the current time of day.

Fig.1 Use-Case Diagram

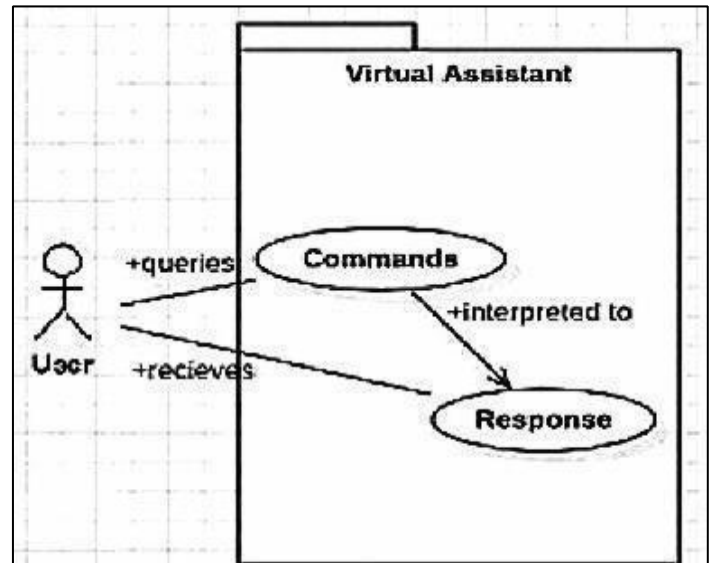


Fig.2 Data Flow Diagram

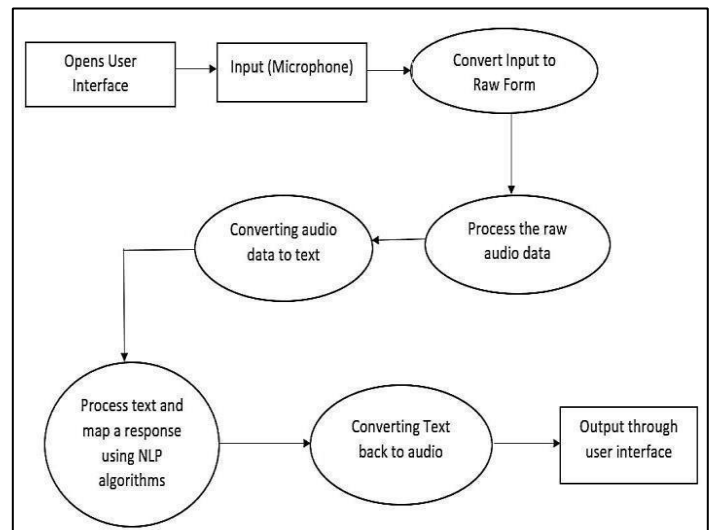
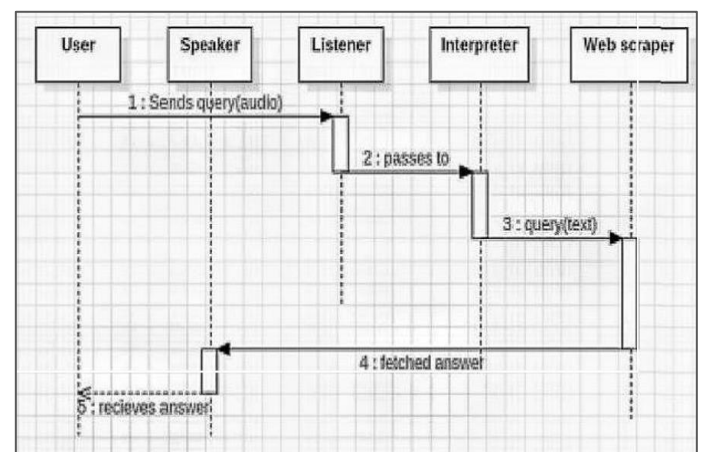


Fig.3 Sequence Diagram



## 5. FUTURE WORK

As AI continues to evolve and gain popularity for daily tasks, its future scope is vast. An upcoming enhancement will be the addition of multilingual support. This feature will enable users to issue commands in Hindi, and the assistant will respond in both Hindi and English. Furthermore, additional features will be introduced to enhance user experience and streamline tasks. The convergence of ancient wisdom with modern technology has the potential to profoundly reshape healthcare practices.

## 6. CONCLUSION

The paper titled "NLP-based AI Voice Assistant" provides a detailed examination of the design and creation process of a voice-enabled personal assistant tailored for desktop use. The project's modular structure enhances flexibility, allowing for the seamless incorporation of additional features without disrupting existing system functionalities.

In today's fast-paced lifestyle, this voice-enabled desktop assistant proves to be exceptionally time-saving compared to earlier methods. Designed with user-friendliness as its primary focus, the assistant efficiently carries out assigned tasks. It not only executes commands issued by the user but also responds to inquiries and statements, such as initiating operations and tasks. The assistant introduces itself to users in a manner that fosters a sense of comfort and encourages open interaction. The program aims to eliminate any unnecessary manual effort required to complete tasks, operating entirely on verbal input instead of traditional physical inputs. Additionally, the assistant is capable of performing a diverse range of tasks, including restarting or shutting down the PC with a simple voice command.

## REFERENCES

- [1] S. Subhash, P. N. Srivatsa, S. Siddesh, A. Ullas and B. Santhosh, "Artificial Intelligence-based Voice Assistant," 2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4), 2020, pp. 593-596, doi: 10.1109/WorldS450073.2020.9210344.
- [2] Felix, S. M., Kumar, S., & Veeramuthu, A. (2018). A Smart Personal AI Assistant for Visually Impaired People. 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI). doi:10.1109/icoei.2018.8553750
- [3] Ee. P. M., Santra, S., Bhowmick, S., Paul, A., Chatterjee, P., & Deyasi, A. (2018). Development of GUI for Text-to-Speech Recognition using Natural Language Processing. 2018 2nd International Conference on Electronics, Materials Engineering & Nano-Technology (IEMENTech). doi: 10.1109/iementech.2018.8465238
- [4] Iannizzotto, G., Bello, L. L, Nucita, A, & Grasso, G. M. (2018). A Vision and Speech Enabled, Customizable, Virtual Assistant for Smart Environments. 2018 11th International Conference on Human System Interaction (HSI).doi:10.1109/hsi.2018.8431232
- [5] Kepuska, V, & Bohouta, G. (2018). Next-generation of virtual personal assistants (Microsoft Cortana, Apple Siri, Amazon Alexa and Google Home). 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC). doi:10.1109/ccwc.2018.8301638
- [6] J. Sayed ,M. W. Ashour, 2022 International Conference on Edge Computing and Applications (ICECAA), DOI: 10.1049/icp.2022.0334
- [7] Vindhya Vasini K; Tejasvi K; Pravallika K , 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS) ,DOI: 10.1109/ICSCDS53736.2022.9760998
- [8] Satyendra Praneel Reddy Karri; B. Santhosh Kumar, Deep Learning Techniques for Implementation of Chatbots, 2020 International Conference on Computer Communication and Informatics (ICCCI) , DOI: 10.1109/ICCCI48352.2020.9104143
- [9] Aashdeep Singh, P. Murugeswari, S D Prabu Ragavendiran, Amanpreet Kaur, AI-based Chatbot for Physically Challenged People, 2022 International Conference on Edge Computing and Applications (ICECAA),DOI:10.1109/ICECAA55415.2022.9936585
- [10] Rani P. J, Bakthakumar, J, Kumaar, B. P, Kumaar, U. P, & Kumar, S. (2017). Voice controlled home automation system using Natural Language Processing (NLP) and Internet of Things (IoT). 2017 Third International Conference on Science Technology Engineering & Management (ICONSTEM). doi:10.1109/iconstem.2017.8261311