



# An AI based Voice Assistant using Speech Recognition

<sup>1</sup>Arul Natarajan, <sup>2</sup>Vidya Rajasekaran, <sup>3</sup>Shaik Mujahid , <sup>4</sup>Amjan Shaik, <sup>5</sup>J Ramesh Babu,

<sup>1</sup>Assistant Professor, Department of Computer Science and Technology, St Peter's Engineering College, Hyderabad, India

<sup>2</sup>Research Scholar, Department of Information Technology, B.S Abdur Rahman Crescent Institute of Science and Technology, Chennai, India

<sup>3</sup>Student, Department of Computer Science and Technology, St Peter's Engineering College, Hyderabad, India

<sup>4</sup>Professor and Head , Computer Science and Technology Department, St Peter's Engineering College, Hyderabad, India

<sup>5</sup>Assistant Professor, Department of Computer Science and Technology, St Peter's Engineering College, Hyderabad, India

<sup>1</sup>[arulthala82@gmail.com](mailto:arulthala82@gmail.com), <sup>2</sup>[vidyarajesh23@gamil.com](mailto:vidyarajesh23@gamil.com), <sup>3</sup>[shaikmujahid518@gmail.com](mailto:shaikmujahid518@gmail.com)

**Abstract:** *The main aim of creating this model is to develop a Artificial Intelligence (AI) based smart voice assistant named Jarvis which can assist the user with the convenience to command the computer system through their voice and to perform certain actions. This is a mini voice-based assistant developed using speech recognition techniques. Speech recognition is the processes of converting speech into the text and the model is implemented using Python API. The Virtual assistants are software programs which makes our day-to-days lives easy by executing tasks by taking commands through our voice. Data privacy can be maintained as the system is allowed only for the specific task that is required by the user with authorized permissions. We can perform operations like sending e-mails through voice, performing Google searches, using YouTube for watching videos, opening Wikipedia, and also to perform many other daily entertainment activities like playing Music, videos, and many more through Single voice commands. The use of speech recognition technology highly assists people with visual disabilities or motor impairments, to use technology more easily and effectively. Jarvis represents a forward approach in the field of voice recognition technology and ensures that it continues to improve our lives while maintaining privacy and security.*

**Keywords:** *Speech Recognition, Voice Assistant, Data Privacy.*

## 1. INTRODUCTION:

### 1.1 Motivation

The motivation behind the project of an Artificial Intelligence which used by machines, and it shows us the capability for thinking like human being. In this Computer-System which designed in such a way that were typically required interaction between human. As all we know that Python is Growing in emerging languages so it become easy for creating a script for a Voice Assistant by Python. The instructions were for the assistant it can be handled as per for the requirements of user needs. The Speech Recognition like Alexa, Siri, etc. In Which Python is used for their Application Programming Interface by using Speech Recognition which that allows us to for converting speech into the text. So, I was interested for this task to Creating own assistant for me. And it becomes easier as it. For doing tasks like sending my emails without typing any word in Keyboard, searching anything in Google without opening any browser, and also performing many other things in daily tasks like we can play music, open our favourite IDE with the help of a single voice command. Nowadays, in the advance technologies we can do such perform for any tasks with a same effectiveness like or we can say more effectively we can do all things. By Creating this type project, in this I realized that were concepts of AI in every field was Increasing and decreasing the human effort for saving time.



## 1.2 Problem Statement

- The accuracy of speech recognition technology is crucial for the proper functioning of voice-based command systems. However, current systems still face the challenges with accurately understanding and recognizing the user commands, especially in noisy environments or with diverse accents and languages. Improving the accuracy and recognition capabilities of voice-based command systems is a key problem to be addressed.
- Natural Language Understanding and Contextual Understanding: Voice-based command systems need to go beyond basic speech recognition and understand the meaning and context of user commands. They should be able to interpret complex queries, handle ambiguous requests, and provide appropriate responses. Enhancing natural language understanding and contextual understanding capabilities is essential to make voice-based command systems more intelligent and intuitive.
- Personalization and User Adaptation: Voice-based command systems should be able to adapt and personalize their responses based on individual user preferences, behavior, and past interactions. Customizing the system to each user's specific needs and preferences can significantly enhance user satisfaction and overall system performance.
- Privacy and Security: Voice-based command systems often involve capturing and processing sensitive personal information. Ensuring the privacy and security of user data is a critical concern. Preventing unauthorized access, data breaches, and ensuring transparent data handling practices are important challenges that need to be addressed to build trust and confidence among users.

## 1.3 Scope

In our daily life, we use our computer to do our personal works and do our office works using keyboard and mouse, and many of the people cannot access the computer using the mouse and keyboard, because they are disabled or do not know how to use the computer, so we have created the application which can access the system using the voice commands. The project is to be developed as personal-assistant for the window-based systems, which can also be used in Linux-based systems to control the system over the voice. They are designed for providing a user-friendly interface to carrying out the several of tasks with the certain and well-defined commands. Automating repeated tasks with voice-recognition applications with the human's time and resources and saves time of humans in do the work as compared without the application.

- Make the Program to be learning lot by its own & designed to develop with a new skill in the Application.
- The Application it can be also developed to the android user's for controlling their phones.
- Voice commands can be encrypted for the maintenance of security of the system.
- Search Engine with voice interactions and Reminder and To-Do application.

## 2. LITERATURE REVIEW:

The "Development of cloud-based smart voice assistant using Amazon Web Services" [1] by Chauhan and Arora in 2023 which describes the development process of a smart voice assistant using an Amazon Web Services (AWS) cloud platform and it was presented at 2023. In the 2nd International Conference Applied Artificial Intelligence and Computing (ICAAIC). It is implementation of cloud-based smart voice assistant technologies. We Used such as speech recognition and natural language processing. It designed for understanding the user commands and provides appropriate results which are based on the input. And the highlight key components and architecture of the smart voice assistant system is which it covers such aspects as speech recognition algorithms, language models, and voice response generation. In AWS cloud infrastructure development, the voice assistant has, including scalability, reliability, and ease of deployment. The "Performance of smart personal assistant applications based on speech recognition technology by using IoT-based voice commands", [2] by Isyanto, Arifin, and Suryanegara which was presented at 2020 in (ICTC). Its performance like smart personal assistant applications by using speech recognition technology and IoT-based voice commands. It focuses on evaluating the accuracy and response time of the speech recognition system when processing voice commands issued through IoT devices. The smart personal assistant applications and the integration of IoT devices for voice command input to measure the accuracy of speech recognition and analyze the response time of the system.

The usability and effectiveness of smart personal assistants. Its findings provide insights into the performance capabilities of the speech recognition technology and its suitability for IoT-based voice command applications. The highlights and potential benefits utilizing this technology offer implications for future research and development in the



field. [3]The "Voice-controlled face recognition based on smart glasses" by Yang and Serikawa in (2020) which was presented to the 8th IIAE International Conference on Industrial Application Engineering in 2020. The Development of a voice-controlled face recognition system which uses the smart glasses. It focuses on the integration of voice commands with face recognition technology for providing a hands-free and convenient user experience. The architecture & implementation of the system, is highlighting the uses of smart glasses as a primary input device for capturing images and processing voice commands. It explains the algorithms and techniques used for face detection and recognition, integration of voice control functionalities. The study demonstrates the feasibility and effectiveness of voice-controlled face recognition based on smart glasses, showcasing its potential applications in various industrial and practical scenarios. The authors present experimental results to evaluate the performance of the system in terms of accuracy and response time. It shows the potential benefits of such a system in industrial and contributes to the advancement of human-computer interaction technologies. [4] The "Voice pathology assessment based on automatic speech recognition using Amazigh digits" by Zealouk, Satori, Hamidi, and Satori in 2018 and it presented in 2nd International Conference on Smart Digital Environment. The Use of automatic speech recognition (ASR) technology for assessing voice pathology in Specifically it focuses on the assessment of voice disorders using Amazigh digits by numeric words in the Amazigh language.

The development of an ASR system is capable of recognizing and assessing the pronunciation of Amazigh digits. The methodology for data collection, pre-processing, and training of the ASR system is developed. It presents the evaluation results of the system's performance in terms of accuracy and effectiveness in assessing voice pathology. [5] The "OCR bases on cheque validation by using image processing" by Kunekar, Vayadande, Kulkarni, Ingale, Kadam, and Inamdar in 2023 was presented in 5th Biennial International Conference in Nascent Technologies Engineering. The development of validation system is based on optical character recognition (OCR) and image processing techniques. It focuses on automating the processing of cheque validation, which involves in manual inspection and verification. The methodology and algorithms are used for cheque image pre-processing, including the techniques for noise removal, image enhancement, and segmentation of key regions. It explains the OCR technique for character recognition, included the features of extraction and classifications methods. [6] The "Beyond Siri and Alexa: Gender and AI Policy" Created by Venugopal and Rituraj in 2022. It explores the intersection of gender and (AI) policy, it specifically focusing in virtual voice assistants such as like Siri and Alexa. The gender implications and biases present in virtual voice assistants which is often portrayed as a female by default. It examines the cultural factors that which contribute by gendered representation The policy dimensions of gender and AI is regulatory frameworks and guidelines that development and deployment of virtual voice assistants.

The importance of considering diverse perspectives and voices in AI policy formulation is ensuring an equitable and representative development of virtual voice assistants. [7] The "Google assistant-controlled home automation with voice recognition" by Yadav, Kumar, Kumar, and Singh in 2021 and presented in 3rd International Conference on Advances in the Computing, Communication Control and Networking (ICAC3N). The development of home automation systems by using Google Assistant with voice-controlled operations is available. The Architecture and components of the system is including the hardware and Software devices, modules involve in automation process. The Google Assistant with a home automation system it explains how the commands are processed and translate into controlling signals for various home appliances and devices. It highlights the uses of voice recognition techniques for enhancing the system's assurance and reliability in understanding user commands for Processing. [8] "Hey, Siri", "OK, Google", "Alexa". Like Acceptance-Relevant Factors of Virtual Voice-Assistants" by Burbach, Halbach, Plettenberg, Nakayama, Ziefle, and Calero Valdez in 2019 it presented in IEEE International Professional Communication Conference. The Work of its acceptance factors associated with virtual voice assistants such as Like Siri, Google Assistant, and Alexa. It focuses to understanding the users' needs and concerns related to the using of voice assistants. The mixed-method approaches, combining the qualitative interviews and quantitative surveys, for gathering insights from users. It explores various factors which influence the acceptance and usage of virtual voice assistants, it including ease of use, privacy concerns, trust, and satisfaction. [9] The "Next-generation of virtual personal assistants by Kepuska and Bohouta in 2018. It presented at the 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC). The next generation of virtual personal assistants is where, specifically Microsoft Cortana, Apple Siri, Amazon Alexa, and Google Home. It provides the virtual assistants, discussing their features, capabilities, and applications. The strengths and weaknesses of each virtual assistant, considering factors such as natural language processing, speech recognition, voice commands, context is awareness, and with other smart devices and services. The current state of these virtual assistants is discussing potential future developments and enhancements.

The virtual personal assistants have various domains, in healthcare, education, entertainment, and smart homes. It highlights the benefits and challenges associated with these technologies. [10] Singular Adaptive Multi-Role



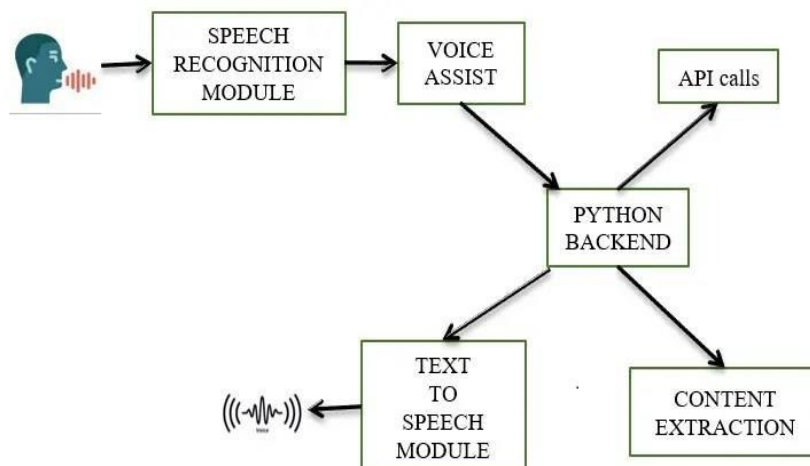
Intelligent Personal Assistant for Human Computer Interaction by Mahmood, Rana, and Raza in 2018. It was presented in 12th International Conference in Open-Source Systems and Technologies. It shows a singular adaptive in multi-role intelligent personal assistant was designed to enhance human-computer interaction. It developed in open-source software framework which combines the various intelligent technologies, in such as natural language processing, speech recognition, and machine learning, for providing a personalized in context-aware assistant. It presents the architecture which consists of multiple modules responsible for different functionalities, including speech recognition, intent recognition, dialogue management, and response generation. It allows learning and adapting to user preferences and behavior over time. They highlight the importance of Personalization in improving user experience and engagement with the assistant. [11]The "The Voice Enabled Personal Assistant for PC using Python" by V., G., C K, G., Kottamasu, M. S., and Kumar, N. P. in 2021 which was published in the International Journal of Engineering and Advanced Technology. It presents a system that enables voice control for personal computers by using the Python programming language. Aims are providing an intuitive and hands-free interaction method for Computers, it allowing users for performing various tasks through voice commands. The architecture and implementation tell the voice-enabled personal assistant. How the system utilizes speech recognition techniques for converting spoken commands into text. The integration of natural language processing and machine learning algorithms to understand and interpret the user's intent from the recognized text is proposed. It supports a range of functionalities, including opening applications, performing system-level operations. [12] Desktop Voice Assistant with Speech Recognition Intelligence Which was published in the International Journal for Research in Engineering Application & Management in 2020. It presents the development and implementation of a desktop voice assistant named as DVAbot. It provides voice-controlled functionality to desktop computers, enabling users to perform various tasks through voice commands. The Architecture of DVAbot, it involves the integration of speech recognition technology and natural language processing techniques. It utilizes automatic speech recognition (ASR) to convert spoken commands into text, and those employs natural language processing algorithms to interpret and execute these commands.

The functionalities of DVAbot include opening applications, performing internet searches, sending emails, managing files, and controlling media playback. The algorithms and techniques used to enable the voice assistant's capabilities. The Evaluation of DVAbot is performing, including metrics such as speech recognition accuracy, response time, and user satisfaction. The conclusion tells DVAbot, a desktop voice assistant with speech recognition intelligence and contributes to the field of voice-controlled interfaces for desktop computers and demonstrates the feasibility of implementing such systems. [13] Personalized Desktop App Based Interactive Means of Zira Voice Assistant presented in 2022 at International Interdisciplinary Humanitarian Conference for Sustainability (IIHC). It discusses the development of personalized desktop application that uses the Zira voice assistant. The aim of the research is to create an interactive and user-friendly voice assistant system for desktop computers for individual user preferences and requirements. The architecture and design of the personalized desktop app, which integrates the Zira voice assistant with various features and functionalities. The app allows users to interacting with their desktops using voice commands, enabling tasks like opening applications, performing searches, managing files, and controlling system settings. The importance of Personalization in the voice assistant system, it enables a more and efficient user experience. The techniques and algorithms are employed to personalize. [14] AI Personal Assistant was published in the International Journal of Information Technology and Computer Engineering. The development and implementation as AI personal assistant aim to create an intelligence assistant that can assist the users with various tasks and provide personalized services. The highlighting technologies and algorithms used in its development. The designed to understand natural language commands and responds accordingly, providing information, performing tasks, and interacting with users in a conversational manner which include voice recognition, natural language processing, machine learning, and task automation. [15] The authors aim to create an intelligent assistant that can assist users with various tasks and enhance their productivity. The architecture and designing, highlighting of the technologies and algorithms are used in its development. For understanding user commands, perform actions, and provide relevant information in a user-friendly and efficient manner, which include speech recognition, natural language processing, and task automation. To learn from user interactions and adapt its responses and behavior based on user preferences and patterns.

### **3. PROPOSED SYSTEM**

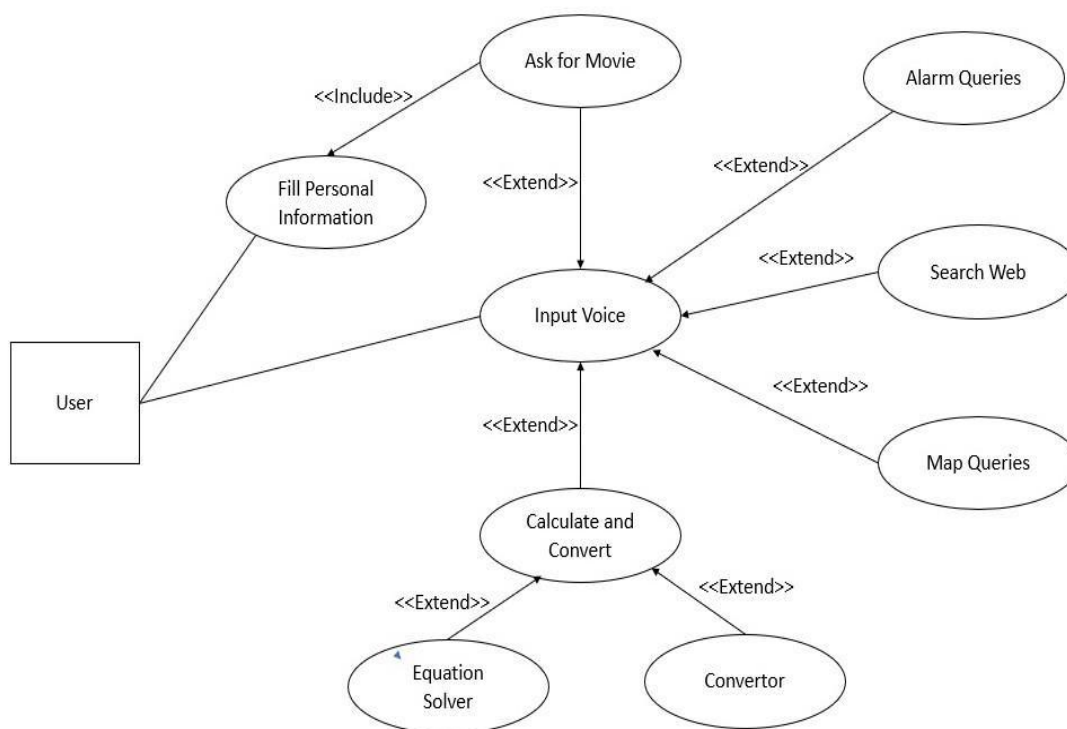
The below representation denotes the architecture block diagram of the proposed system.

#### **3.1 Block Diagram Representation**



**Fig 3.1.1 Architecture Block Diagram**

Fig 3.1.1 talks about the Process were the System Takes the Voice Command by the help of Speech Recognition and Process the Requirements of User. Its calls API function in used VOSK and KALDI for Voice to Text Speech.



**Fig 3.2.2 Use-Case Diagram**

Fig 3.2.2 shows the user calls, the command in system takes the input as voice and connects to its related functions and gives the output of its command.

### 3.2 GUI Representation

Fig 3.2.1 represents the GUI representation of the developed model. This is the initial page setup when the program is executed.

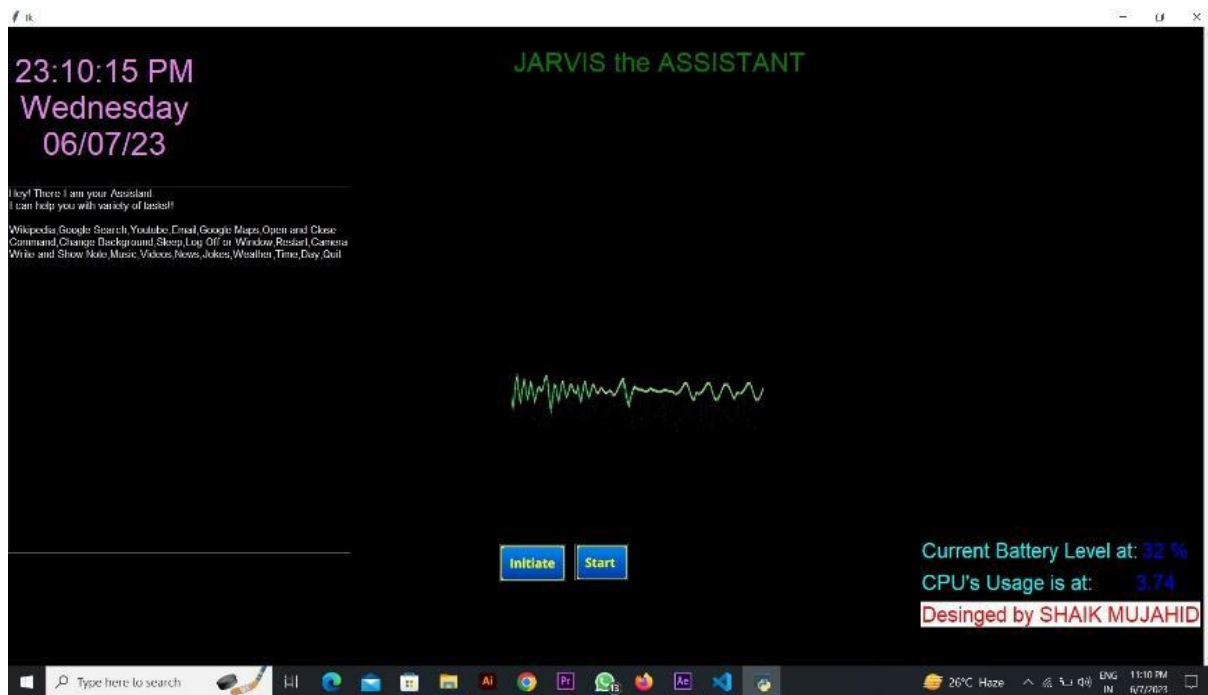


Fig 3.2.1 Screenshot of Proposed System

### 3.3 Software Design

In designing the software following principles are followed:

- **Modularity & Partition:** In Software the designed has each system for to consisting the hierarchy and modules for the partition were it to be separated in different functions.
- **Interacting:** In the modules were should have little dependence on other modules of a system.
- **Cohesion:** In this the modules are being carried out in a single processing function.
- **Sharing:** To avoiding the duplicates and allowing those into a single module are called as other needs to the functions which it provides.

## 4. RESULTS AND DISSCUSION :

### 4.1 Input Design

The Requirements, procedures of Inputs are collecting the necessary inputs in data. The most proficiently designs of input has been designed to be displaying it, the communication of the users with the system has-been most effectively and simplifying way.

To measure those things were taken this following

- It Controls all data in the given inputs.
- Eliminating extra steps
- Keeping the process simple
- The total inputs forms are being viewed the design.
- Surrounding ambiance should be taken care of.

All the inputs are taken are been created for the user's voice commands. It while taking input the interface should be maintained user friendly.

### 4.2 Output Design

The output interface is shown with the fetched results which are requested by the user and it might be opening a website laying music or having a fun chat anything which could not be processed is shown with the error description which makes the user clear it using the menu options it provides.

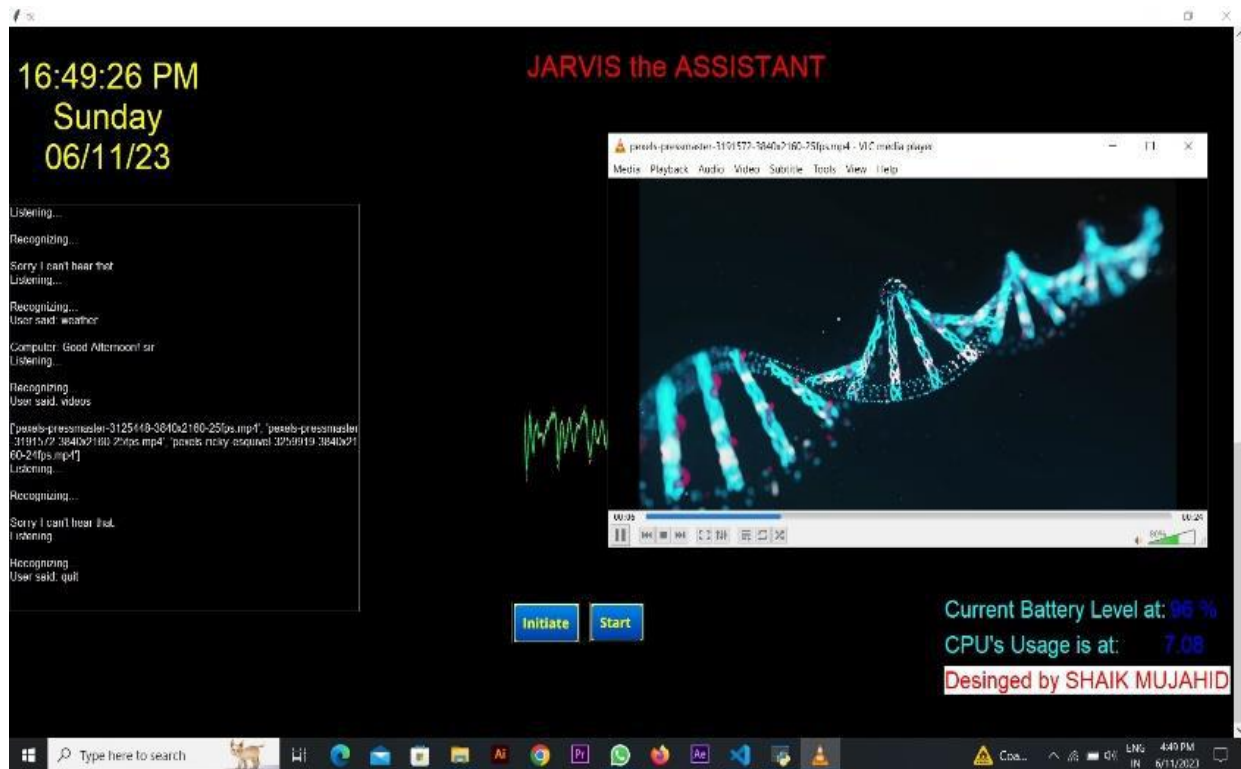


Fig 4.2.1 Screenshot of Video Representation

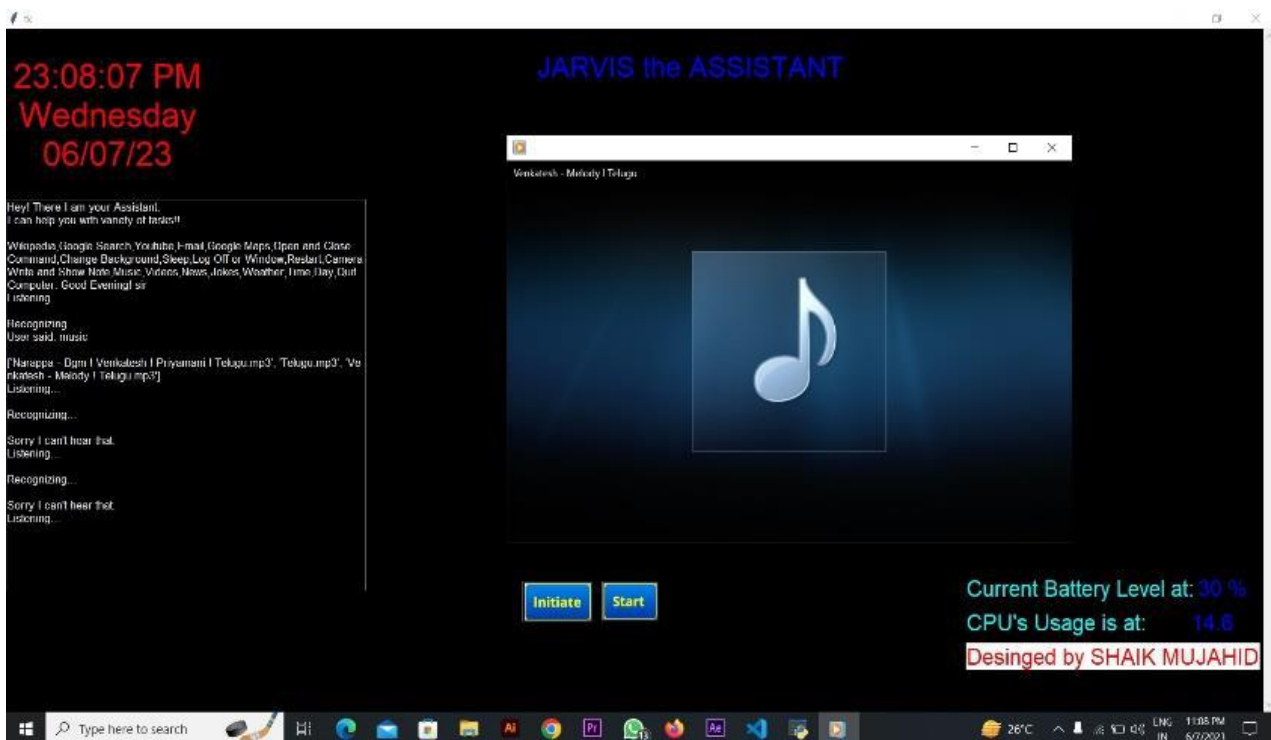
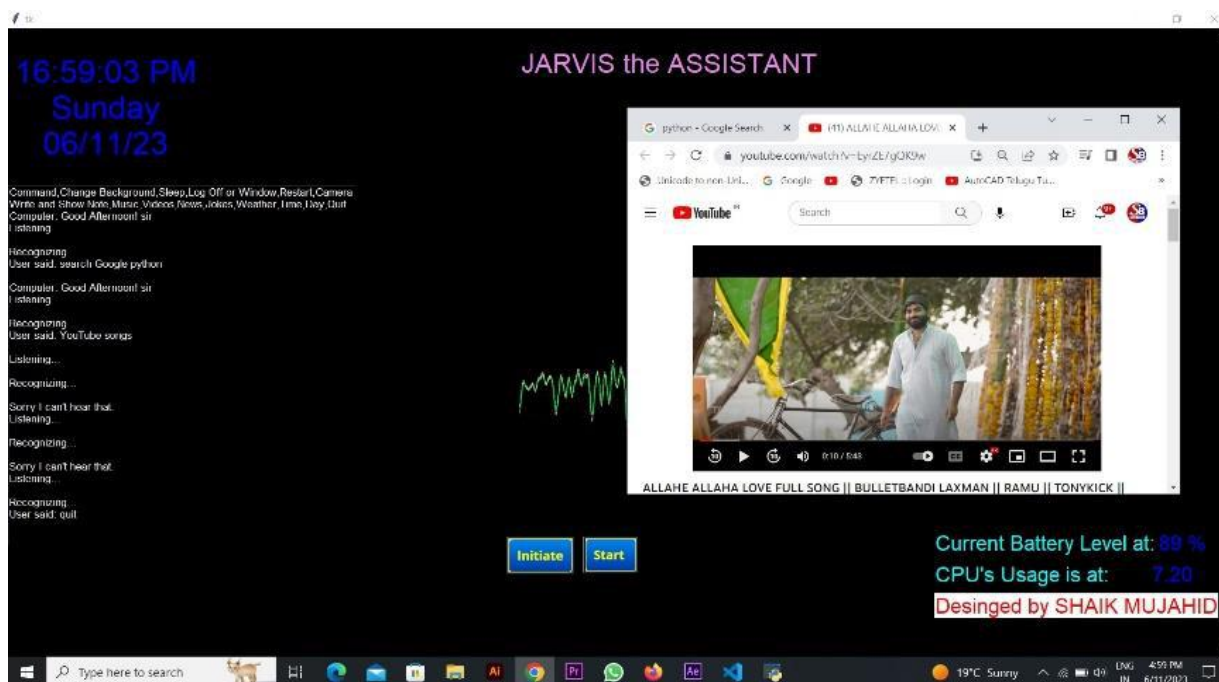


Fig 4.2.2 Screenshot of Music Representation



**Fig 4.2.3 Screenshot of YouTube Representation**



**Fig 4.2.4 Screenshot of Wiki Representation**

## 5. CONCLUSION

We have modeled the Application for the computer using python and methods of while, for loops, functions, and if-else statement. In this project the computer system is being controlled over voice and can convert the speech into the text to take input and to make the computer speak the text is converted into speech/voice. In the future we can add more categories, add Artificial Intelligence in the application and make it more user friendly and improves its quality. I and my team members are interested in studying these methods and implement in another AI and Machine learning. This application is very useful and the smart voice assistant work without any doubts as it helps for saving time and the user has been interacts friendly, it ineffectively and efficiency. While doing this project, I realized that the purpose of advancement in the futures and there were some limitations encountered.





## REFERENCES :

1. Chauhan, S., & Arora, D. (2023). Development of cloud based smart voice assistant using Amazon web services. 2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC). <https://doi.org/10.1109/icaaic56838.2023.10140284>
2. Isyanto, H., Arifin, A. S., & Suryanegara, M. (2020). Performance of smart personal assistant applications based on speech recognition technology using IoT-based voice commands. 2020 International Conference on Information and Communication Technology Convergence (ICTC). <https://doi.org/10.1109/ictc49870.2020.9289160>
3. Wenji, W. (2021). Retracted article: Recognition of rural e-Commerce smart assistant system based on smart voice technology. *International Journal of Speech Technology*, 25(S1), 21-21. <https://doi.org/10.1007/s10772-021-09887-z>
4. Yang, Z., & Serikawa, S. (2020). Voice-controlled face recognition based on smart glasses. *Proceedings of the 8th IIAE International Conference on Industrial Application Engineering 2020*. <https://doi.org/10.12792/iciae2020.056>
5. Zealouk, O., Satori, H., Hamidi, M., & Satori, K. (2018). Voice pathology assessment based on automatic speech recognition using Amazigh digits. *Proceedings of the 2nd International Conference on Smart Digital Environment*. <https://doi.org/10.1145/3289100.3289117>
6. Kunekar, P., Vayadande, K., Kulkarni, O., Ingale, K., Kadam, R., & Inamdar, S. (2023). OCR based cheque validation using image processing. 2023 5th Biennial International Conference on Nascent Technologies in Engineering (ICNTE). <https://doi.org/10.1109/icnte56631.2023.10146687>
7. Venugopal, V., & Rituraj, V. (2022). Beyond Siri and Alexa: Gender and AI policy. <https://doi.org/10.4337/9781800377462.00017>
8. Yadav, S. S., Kumar, P., Kumar, S., & Singh, S. (2021). Google assistant-controlled home automation with voice recognition. 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N). <https://doi.org/10.1109/icac3n53548.2021.9725502>
9. Burbach, L., Halbach, P., Plettenberg, N., Nakayama, J., Ziefle, M., & Calero Valdez, A. (2019). "Hey, Siri", "OK, Google", "Alexa". Acceptance-relevant factors of virtual voice-assistants. 2019 IEEE International Professional Communication Conference (ProComm). <https://doi.org/10.1109/procomm.2019.00025>
10. 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). <https://doi.org/10.1109/ccwc.2018.8301638>
11. Mahmood, K., Rana, T., & Raza, A. R. (2018). Singular adaptive multi-role intelligent personal assistant (SAM-IPA) for human computer interaction. 2018 12th International Conference on Open-Source Systems and Technologies (ICOSST). <https://doi.org/10.1109/icosst.2018.8632189>
12. V., G., C K, G., Kottamasu, M. S., & Kumar, N. P. (2021). The voice enabled personal assistant for PC using Python. *International Journal of Engineering and Advanced Technology*, 10(4), 162-165. <https://doi.org/10.35940/ijeat.d2h4425.0410421>
13. Desktop voice assistant with speech recognition intelligence (DVAbot). (2020). *International Journal for Research in Engineering Application & Management*, 301-305. <https://doi.org/10.35291/2454-9150.2020.0411>
14. Joshi, A., & Akram, S. V. (2022). Personalized desktop app based interactive means of Zira voice assistant. 2022 International Interdisciplinary Humanitarian Conference for Sustainability (IIHC). <https://doi.org/10.1109/iihc55949.2022.10060610>
15. Kumar, P., Mhaske, P., Mali, R., & Shukla, P. (2023). Jarvis the AI personal assistant. *International Journal of Information Technology and Computer Engineering*, (32), 16-20. <https://doi.org/10.55529/ijitc.32.16.20>
16. Shashwat Shukla., Srishti Sinha., S. S. (2021). Jarvis: Desktop assistant. *International Journal for Modern Trends in Science and Technology*, 7(05), 178-183. <https://doi.org/10.46501/ijmtst0705030>
17. Henley, K. (2022). Hey, Siri, am I sexist?: An examination of AI voice assistant gender on perceptions of helpfulness and usefulness. <https://doi.org/10.31219/osf.io/4whrx>