A Machine Learning Based Approach to Control Device using Voice Recognition System

[1]Rahul Baruah, [*]Sonali Mondal, Biswajit Das

Department of Computer Science, Arunachal University of Studies, Namsai, Arunachal Pradesh, India

Abstract: The introduction and extensive use of the Internet of Things, artificial intelligence technologies are beginning to actively used in human life. One of the most significant developments in artificial intelligence is the ability to understand the natural language spoken by the humans. Voice assistance this technology helps in controlling our new generation smart phone or our personal desktop such as laptops and computers. We speak our desire commands to ask queries such as opening applications, sending our texts, or searching our queries on the internet. There are more advance devices that efficiently allows us control the hardware settings that are available in our computer such as increasing volume or resetting screen brightness, using our voice commands. Voice assistant system also controls smart home such as google assistant in our smart phone, Amazon Alexa, or the famous Apple Siri, and also, we can control our home appliances such as our thermostat, door locks or the lights of our home, using just our voice. This paper consists a comprehensive design and of mobile automation and high-end security system using artificial neural networks. For this program that is the voice recognition system, one of the famous programming languages is used that is the Python. HTML, CSS, and JavaScript is used to give a distinctive user interface in a form of web application for the voice assistance system

Keywords - **Artificial Neural Network, Natural Language Processing, Machine Learning, Voice detection, Text to speech, Voice recognition.**

I. INTRODUCTION

In our technological age that is the twenty-first century, machines are replacing humans in very capacity. In the past years there have an incredible growth to automation or we can say in new generation technologies. We know we give instruction to our computer in the input device such as keyboard and mouse, just imagine a computer does all our task just by listening to our voice but yes, the science is so advance now a days that it is possible. To achieve this, we must use an API which compiles our human spoken languages or the voice messages into text messages to give the input to the computer. There are many Multinational Companies like Google, Amazon and Apple are doing their best to achieve this in a very universal manner. As we know voice assistant has been very crucial role to the technological development and there are promising future capabilities such noise reduction and a high-end security system. Controlling a device using voice recognition, it is a program fully developed by python that helps us controlling a device for example, sending a WhatsApp text to someone asking the weather, asking to call some, asking the date and time, asking about, and everything that a device does with the help of our voice. The voice recognition system uses the modules like speech-to-text, and a language processing algorithm to perform all the commands or task by recognizing the voice of the user. As per the user's requirement the assistant filters out the unwanted noises and carries out the given task.

Understanding the importance of all this, I have created a rather interesting and helpful program that can installed anywhere and can be a helpful assistant to anyone who can speak. It is very helpful for differently abled person; this program is rather made for the people who wants to achieve many things in their life but cannot do anything because of the physical problems of their bodies. This program will help them in doing task in their computer system which they were unable to do previously. As Duta and et al stated that the rapid advancement of technology comes from the development of innovative solutions to control devices using voice recognition systems. These systems use machine learning algorithms to recognize and interpret voice commands, providing users with a convenient, hands-free way to control their devices [1]. The prevalence of machine learning-based voice recognition systems in everyday life has become increasingly widespread, with applications ranging from virtual assistants on smartphones to voice-controlled home automation systems. These advancements in technology have simplified human-machine interactions and made device control more convenient [2]. Speech recognition is a relatively new addition to the virtual world. Though it is somewhat successful, it is not very helpful and is not used by the user due to the high error rate. The future virtual assistant is not yet ready to become a regular part of the user's life, even though its error rate is under 10%. The project's objective is to create a virtual assistant with low error rate speech recognition as a consequence. I created a voice assistant that reduces the number of input devices by enabling users to complete any job on the system without using a keyboard. These kinds of visual aids are extremely useful for elderly people, blind &/or visually impaired people, kids, etc. by making it easier for people to interact with the machine. Even the blind people who can't see a machine can communicate with is using their voice.

Some of the crucial feature in the system that I have developed can do is Reading newspaper, getting updates on email, searching on the web playing music or video online, Running any programmer application in the system, the hot word detection, ChatGPT API Alternative, Chat History feature, WhatsApp automation and mobile automation.

II. LITERATURE REVIEW

Chinchane and et al stated in their research paper that technology has advanced to the point that many elements of our life, including device control, are now more straightforward and efficient. These days, speech recognition technology makes it simple for consumers to control gadgets with just their voice. This speech recognition-based machine learning method of controlling devices have several advantages and possible uses [2]. Mokgonyane, et all stated in their research paper that through python and also speech recognition module that comes with python that it pyttsx3 that automatically connects the voice recognition and the device. It allows the user to give command in the form of voice to the computer for example "Open Chrome Browser", the system than analyze the command and then open Google Chrome browser. It also controls other devices using the voice command that are very easy to understand and follow. The voice recognition technology differentiates between infinite words and also the model data in the different types of ways that are paved for controlling different device [3]. Shakil and et all stated in their research paper that the most important thing about this research is converting the audio input into text using python diabetes and python modules. The Machine learning algorithms are used to convert or to extract the future that are relevant in analyzing the audio data [4]. Chinchane and et all said that the understanding the basic principle of voice recognition and how the device is controlled by using the voice recognition the software developers can Increase the UI or UX or the user interface and improve the user efficiency and also be able to control the device without using any input device or hands thus only by our voice commands. But also using voice recognition for controlling the device also increases the difficulties for security reasons we have to come with new methods and explore a very secure authentication to prevent from unauthorized access or the hackers to steal anything or to change the confidentiality of the system [1].

Ali and et all in stated in their paper that it is very important to be sure about the security and the verification of a user and also deny the unauthorized access to other users who are not using the system. So, a better practice should be used to ensure the privacy of a user and protecting the confidentiality of the system. The voice recognition System has developed so much in the 21st century that it helps a user to operate any device that are given in the common line of the coding of the particular device to be controlled by using easy human spoken words or commands. The voice recognition systems offer to be easy, to be more enjoying, it is more approachable and it also improves the efficiency in controlling a device using voice command. The voice recognition system has also been developed to be expanded in the potentiality of applications regarding the device control which also allows the interaction between the user and the device and also completing all the tags without any need of using a keyboard or a mouse or any other input device. The voice recognition has also removed the basic need of the manual input and made the controlling of devices easier and more accessible and also more convenient [5]. Tandel and et all stated in their paper that peoples that are differently abled such as the blind, can also take the advantage of the technological advancement. Voice recognition system has that kind potential that can transfigure the device for simple task just by using their voice, thus voice recognition has too many merits and also has application that have potential enough to help them [6].

III. EXISTING WORK SURVEY

Table-1: Existing work survey

Works	Name of the System	Technique that has been used	Type of the System	Research about the System
Ayush Chinchane and et all [2]	SARA: A Voice Assistant	Python, Machine learning and Natural Language Processing	Voice Assistant	The study discusses creating a voice assistant system, having a focus on SARA, a focused-on AI voice assistant. Python is used as the primary programming language in the system's design for a number of activities, like text-to-speech, voice command, and speech recognition. The voice assistant is designed to help users with many tasks such as reading emails, doing online searches, and playing music.
Lalit Kumar [8]	Desktop Voice Assistant Using Natural Language Processing	Speech Recognition, Python, system Calls, Google-text-to-Speech	Desktop Voice Assistant	In this research paper, a voice assistant has been developed that carries out all the task given by the users. In this project a special feature is added the it only listens to the user's voice
Jayaram and et all [9]	Desktop Assistant AI Using Python	Text to Speech, Speech to Text, language Processing, Voice recognition, Pyttsx3, Speech Recognition, SQLite	Desktop Voice Assistant	In this project a python-based voice activated personal assistant is discussed. This system works online, plays music, search in Wikipedia and also open some desktop application. It works online only
Vora and et all [10]	Jarvis: A PC Voice Assistant	Python and Google Text to Speech	PC Voice Assistant	In this project the voice assistant has many types of automation using single line command. It reduces the work effort and makes it easy to do task like searching on the Web, weather forecast, Language translation.
Patil and et al [11]	Intelligent Voice Assistant	Speech recognition, Python, and google Text-to Speech,	Voice Assistant	In this paper, the assistant gives back the response to every command given by the users. The voice recognition system has to wake before doing its given task. Here, various skills are created in Hindi and Marathi languages, such as facts, weather, time. Nearby hospitals, and

		API-Application		city guide. These types of skills are developed through Amazon Developer
M.J. and et all [12]	A voice-based text mail system for visually impaired	Voice-based programming and visually handicapped, Email system	Voice-based Text Mail System	In this project an android application has been proposed specially for visually challenged people. A voice-based mailing service has been used where mails can be sent and read by them on their own. Here users use certain keywords to perform specific actions like composing mail etc.
P.M. Dias and et all [13]	Virtual Assistant in Native Language	Virtual Assistant, Sinhala, Cloud Deployment, Speech recognition, Translation	Voice-Assistant	In this paper the Sinhala language is used. The tech industries today use this language in major real time project. This project certainly tells us how a virtual assistant works.
P. Sing and et all [14]	Voice Control Device using Raspberry Pi	Personal Assistant, NLP, Raspberry Pi, Querry processing	Voice-Controlled Device	In this paper it is stated that the voice command is implemented in such a way that it is intelligent system. Here he Raspberry Pi is the primary hardware for the voice assistant. The text to speech by using simple text then the query processing and NLP is used to interpret the main meaning of the command given by the user.
Aktar and et all [15]	Voice Recognition based Intelligent Wheelchair and GPS Tracking System	Voice Recognition, GPS module, smartphone application, obstacle	Intelligent wheel and GPS tracking system	In this paper a voice recognition based intelligent wheelchair system for differently abled people who are unable to drive through hand. This System work in a different type the differently abled who can use wheelchair through voice, the location of the person can be tracked by GPS module that is used in the module in the wheelchair and that function gives the information or the details in the smartphone.

IV. PROBLEM STATEMENT

The security concerns are one of the major problems I have come through by understanding the previous work that has been done and with the technological advancement, the threats regarding the confidentiality of the users' personal files such as ID card replication, PINs and passwords theft, while using such application. We have done our project by using Artificial Neural Network to reduce the security concern to a extend. While using a voice recognition system there are too much of error rate in speech recognition. As different people have different accent, sometimes it becomes difficult for the voice assistant to understand what exactly the command is. The voice recognition system should work as an AI assistant just like a normal human being talking to another human being. We also faced problems while installing some modules like pyaudio as the current pyaudio module doesn't works directly we have to install an older version of the module. A stable internet is must to use the program. Although the problems that we faced are not so major and we are working on them. Our system doesn't have any drawbacks and works perfectly fine. As I have research, I have not found any assistant with WhatsApp automation and ChatGPT API so I have tried to add this features in the system.

V. PROPOSED SYSTEM

After studying the system, the proposed system focuses on making day to day task a bit easier to a certain extend for example the WhatsApp automation which will help the user to Voice call, send a Text message and even video call. This project is specially for the differently abled peoples who has lost their sight or by any chances lost their hands, this system will thus help them to use a desktop to a very extend and also can enjoy the upcoming technology. There is also a chatting feature in the system where we can with in AI as if it is a real person, it gives a human-to-human interaction to the user. We have also implemented the famous ChatGPT API alternative it tells the user everything the users ask for example if a user ask "Jarvis give me the code to center a div in html" the system will write back the whole code for the user. Then there is the chat history feature in the system that it every command we give to Jarvis and every reply that the system gives us back, the system stores in the chat history feature. The system will have women voice to communicate. If the system cannot recognise what the users is saying or the users' command the system will again come back to the start. This system will have a unique wake command or the Hotword feature as we have to say a particular code word to Jarvis to listen to us and give us our output. This system will be totally be made by using python and it's in built coded environment.
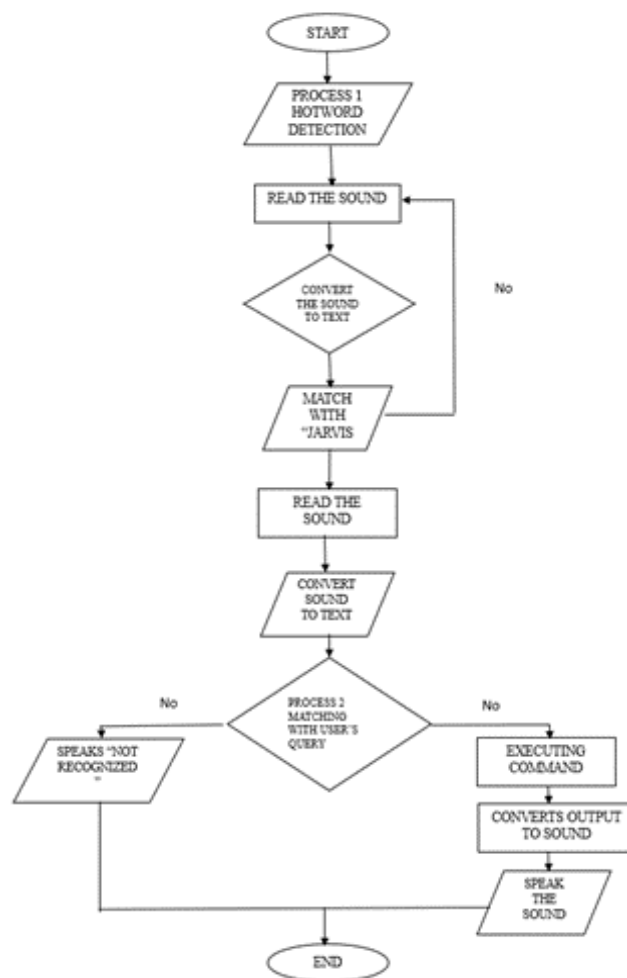
Fig-1: Data Flow Diagram

VI. WORKING OF A VOICE RECOGNITION SYSTEM

The user asks a personal assistant to perform task. Then Natural Language Processing audio signal is converted into digital data that can be analysed by the software. Then data compares the command with the database of the system using an algorithm to find a suitable answer. The database is located on python that is been given in pre-installed. For this very reason a stable Internet connection in needed. The program does all the task that has been set to the voice recognition system, as it does what the users ask it to do. The system then breaks down the audio into individual sounds, then the ADC (Audio-To-Digital-Converter) translates the sound waves to digital data. Then the system uses the algorithm to find the most probable word to fit in that language. The machine text phonemes and connects them to dictionary word.

i. Analog-To-Digital Converter

The ADC removes all the unwanted noise and sound in a sense it reduces the noise reduction. The it also normalizes the sound and speed of speech as different people speak differently.

Fig-2: Analog to Digital Converter Model

ii. Spectrogram

A spectrogram is a graphic display of a signal's frequency range that changes over time. In signal processing and analysis, it is used to explain the frequency content of a that change over time signal. A spectrogram gives a changing view of how the frequency content is distinct over time, compared to a simple frequency plot, which shows the frequency content at only one point in time. Here the darker area shows the high intensity of the sound wave and the lighter area indicates the low density of the sound wave.
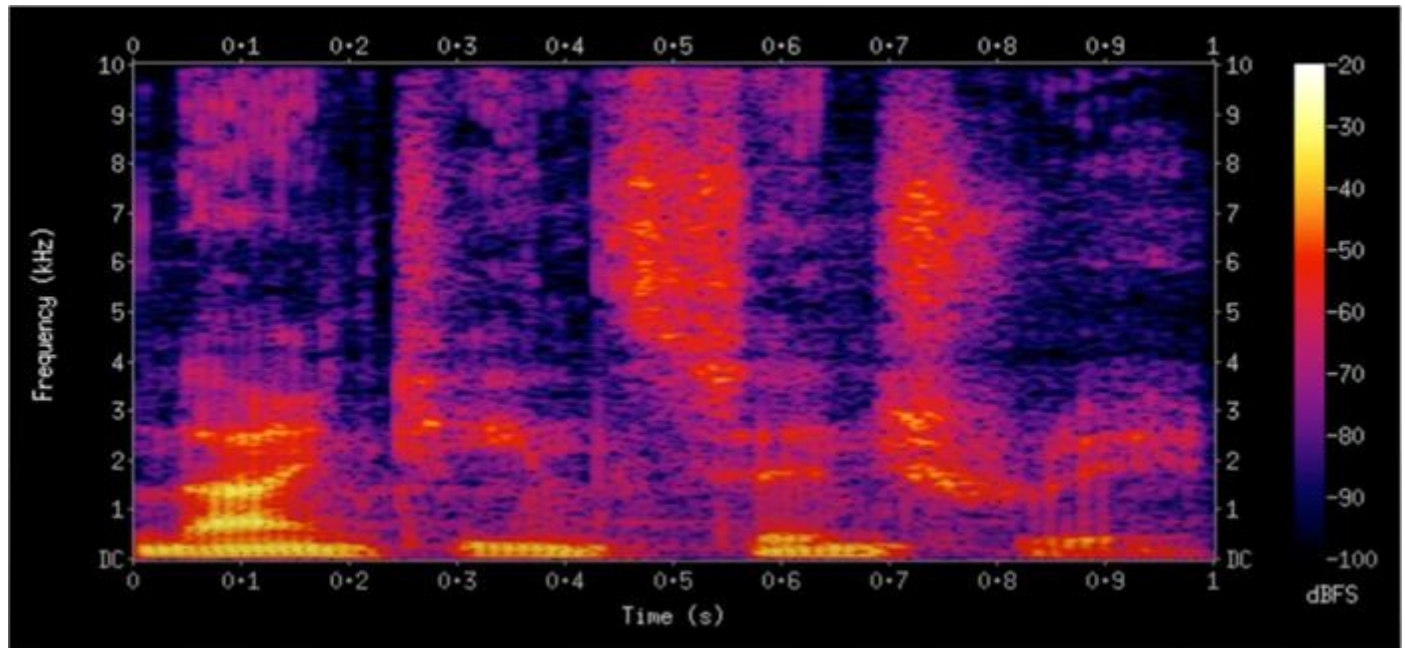
Fig-3: Spectrogram

VII. NATURAL LANGUAGE PROCESSING

Natural Language Processing enables a machine to take what we are saying, make sense out of it, and formulate a statement of its own. The NLP allows the voice recognition systems to convert the real intention of the user's voice and understand the meaning and then gives out the output accordingly. The famous Voise assistant like the Amazon's Alexa and the infamous Google Assistant, this voice assistant system uses the NLP technique to understand the voice commands and response to the commands accordingly.

It uses variety of technique to accomplish this, and they are:
i.   Segmentation
ii.   Tokenizing
iii.   Stop words
iv.   Stemming
v.   Lemmatization
vi.   Speech Tagging
vii.   Name Entity Tagging

i.   Segmentation: Segmentation is one of the necessary steps the NPL for the voice recognition system. The segmentation divides the flow of speech into smaller and a meaning segments or words

ii.   Tokenizing: Tokenizing is one of the important stages in voice recognition system. Tokenizing breaks down human spoken language into small parts known as token, it includes words grammatical information or grammatical errors and punctuation.

iii.   Stop Words: Stop words are basically words which exclude attribute extraction as they don't carry any meaning. E.g. "the", "and", "is" etc.

iv.   Stemming: Stemming is a method applied in Natural Language Processing (NLP) to cuts down the words to their base form. In Voice Recognition System, it helps in enhancing accuracy voice recognition cutting down words to their original form.

v.   Lemmatization: Lemmatization is the technique of cutting down words to its basic form called lemma. Its alike stemming, but in lemmatization meaning of words is determined based on part of voice.

vi.   Speech Tagging: Speech tagging is called part of speech (PoS¬) tagging, it's the method of allocating grammatical stamp to every word in the provided voice, premised on usage. In term of voice recognition system, voice tagging helps enhancing precision of converting speech to text.

vii.   Name Entity Tagging: Name entity tagging (NET) is the process of locating and classifying named entities in text such as people location and dates it is referred to as named entity tagging (NET) by finding the most likely named entities in the used words net can help in improving the accuracy of speech to text conversion in the setting of voice recognition
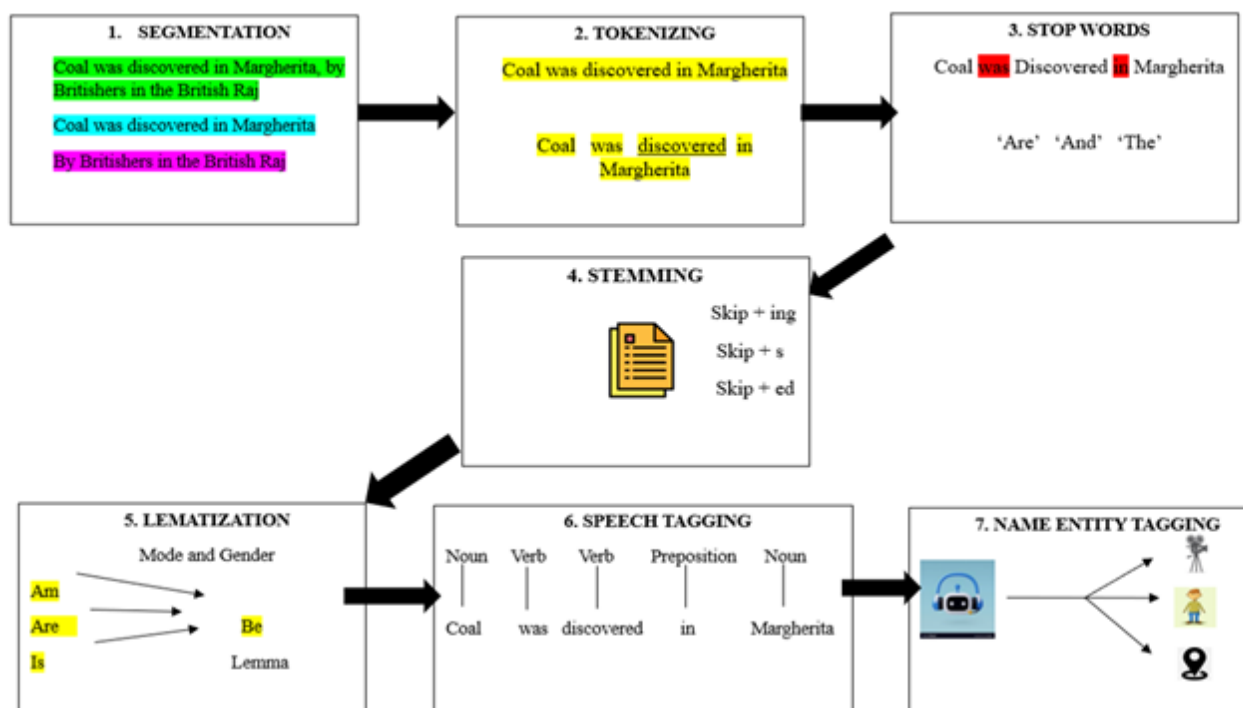
Fig-4: The Natural Language Processing Stages

VIII. ARTIFICIAL NEURAL NETWORK

The power of artificial neural networks (ANNs) to learn complex shapes and recreate non-linear interaction results in their extensive use in systems that recognize speech. ANNs can be used in the language modelling, feature extraction, and classification stages of a voice recognition system. ANNs can be used to extract specific characteristics from unprocessed audio signals in the feature extraction step. For example, an artificial neural network (ANN) that duplicates the human auditory system can be used to extract Mel-frequency Cepstral Coefficients (MFCCs). ANNs can be used to map the obtained features to the connected phonemes or words in the classification stage. As they can imitate time relations in speech signals, recurring neural networks (RNNs) and long short-term memory (LSTM) networks are frequently utilized for this purpose.
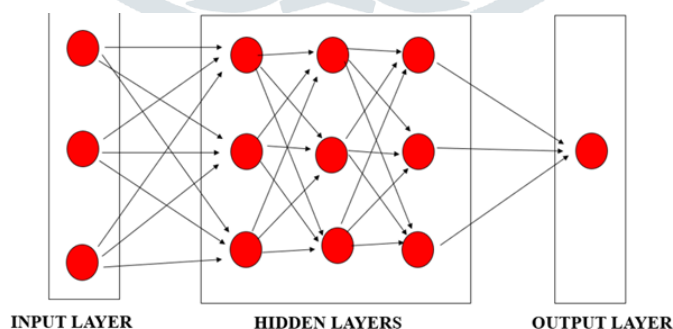

Fig-5: The layers in Artificial Neural Network

IX. WORKFLOW AND METHODOLOGY

The project begins with process which is Hotword detection, it's like a wakeup call to the system, which acts as a command that goes through the microphone. Then the users give a command which includes getting any information, or accessing the computers data and many more. Based on reading the data stated above and verifying the cases, this is actual research. The python programming tests the data from the datasets and other internet resources.
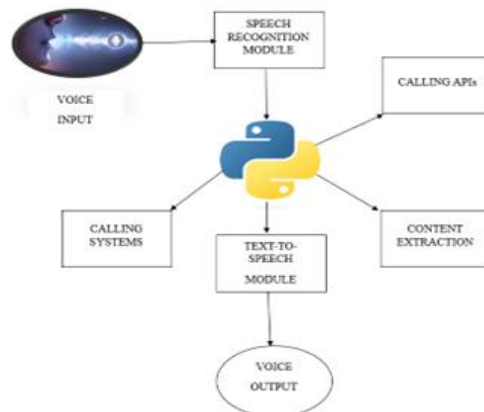
Fig-6: The Workflow of the system

The Speech Recognition Module converts the vocals or the voice command into text using Google's online speech recognition system. The voice file texts from specific corpora are open for users on the information centers computer network server. They are only had on the system before being sent to Google Cloud for speech recognition. Then the Python Backend interprets all modules of voice recognition system and identify if the input or the command is API call or system call. The output is then returning back to the python backend and gives out the desired output to the user. Calling API's, it is an Application Programming Interface is software interface that helps two applications to connect to each other. Content Extraction it is a process extracting organised data from unorganised data or half organised document that document can understand automatically is called content extracting. Calling System is a method which is programmatic where a computer program requests a service from kernel of the installed operating system. Process scheduling is executed in the kernel services. Then the Text-to-Speech This module enables the computer to read text via sound or we can say the computer talk back. The written text is converted to audio file that is then represented in audio waveform as the sound. User Interface gives out graphical interface to the user, the python backend and frontend is connected with the eel module.
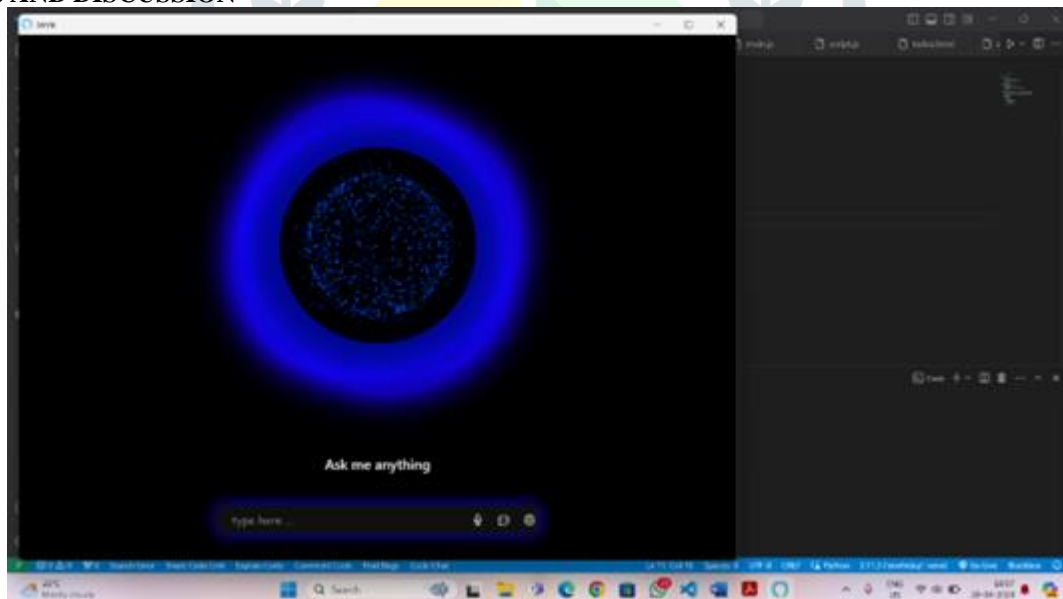
X. RESULTS AND DISCUSSION

Fig-7: The Landing Page

In this part of the paper, it describes about the output of the project. In this project we have chosen python as programming language. We mainly used AI and machine learning. The main reason of using python is it is easily deployable in Visual Studio Code, as it the most comfortable and easy to use IDE.

The first thing we have to do is to speak out the hot word or the wake-up word for the system that will be "Jarvis" after that we are open to give any command through voice.

This are some of the screenshot and command of systems output:

1. "Jarvis! Tell me about yourself" this is the chatting feature of the system.
2. "Jarvis! Give me a C++ code to insert an element in a linked list." This is the ChatGPT Alternative API feature.
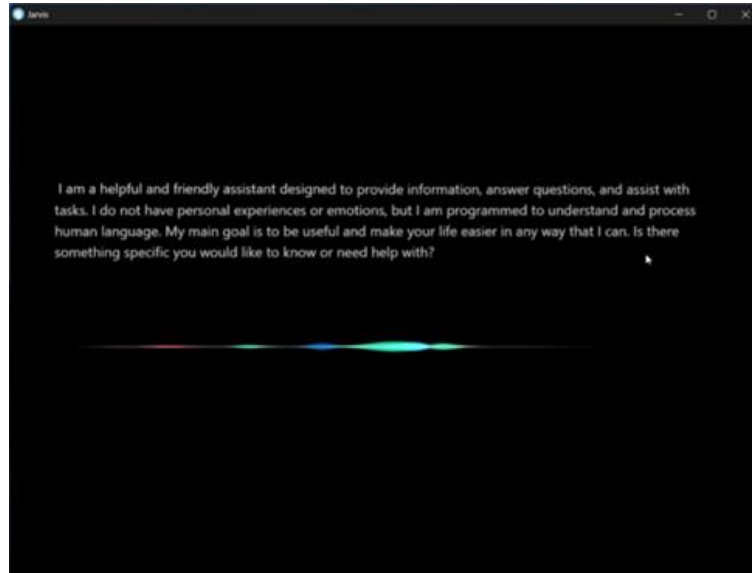3. And the last screenshot is showing the chat history feature of the system.

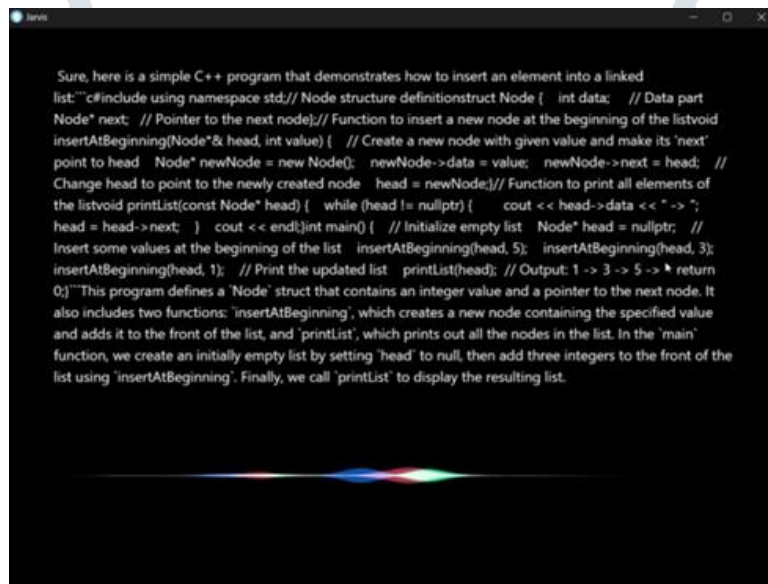**Fig-8:** The Chatting Feature of the system

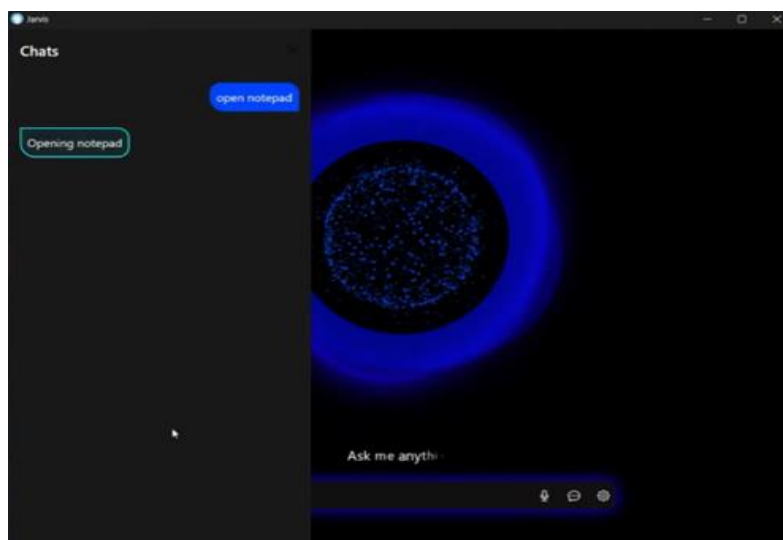**Fig-9:** The ChatGPT Alternative API

Fig-10: The Chat History Feature

XI. CONCLUSION AND FUTURE SCOPE

This Voice Recognition System that I have built can do almost everything that the users ask the system to do from chatting with the system to asking any question in the ChatGPT API and also the calling, sending text in the WhatsApp automation. We have used important packages for python which are speech recognition, Pyaudio pyttxs3 and with the help of this we have made a working project that can wake up with the word "JARVIS" and then with the commands like opening YouTube, open chrome browser, playing music etc and then it gives out the exact result without any error.

For the future I have left some work so that someone can implement features as making the User Interface more attractive, and making the security of the system more secure and also increase the use of machine learning and artificial intelligence and Neural Network and many more things that comes with it and if we achieve this, we can increase the compatibility of voice recognition system.

REFERENCES

[1] Duta, N., 2014. Natural language understanding and prediction: from formal grammars to large scale machine learning. Fundamental Informaticae, 131(3-4), pp.425-440. Doi: https://doi.org/10.3233/FI-2014-1023

[2] Chinchane, A., Bhushan, A., Helonde, A. and Bidua, K., SARA: A Voice Assistant Using Python. International Journal for Research in Applied Science and Engineering Technology, 10(6), pp.3567-3582.. Doi: https://doi.org/10.22214/ijraset.2022.44517

[3] Mokgonyane, T.B., Sefara, T.J., Modipa, T.I., Mogale, M.M., Manamela, M.J. and Manamela, P.J., 2019, January. Automatic speaker recognition system based on machine learning algorithms. In 2019 Southern African Universities Power Engineering Conference/Robotics and Mechatronics/Pattern Recognition Association of South Africa (SAUPEC/RobMech/PRASA) (pp. 141-146). IEEE., Doi: https://10.1109/RoboMech.2019.8704837.

[4] Shakil, S., Arora, D. and Zaidi, T., 2022. Feature based classification of voice based biometric data through Machine learning algorithm. Materials Today: Proceedings, 51, pp.240-247.https://doi.org/10.1016/j.matpr.2021.05.261

[5] Ali, A.T., Abdullah, H.S. and Fadhil, M.N., 2021. Voice recognition system using machine learning techniques. Materials Today: Proceedings, pp.1-7 Doi: .https://doi.org/10.1016/j.matpr.2021.04.075

[6] Tandel, N.H., Prajapati, H.B. and Dabhi, V.K., 2020, March. Voice recognition and voice comparison using machine learning techniques: A survey. In 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS) (pp. 459-465). IEEE., doi: https://10.1109/ICACCS48705.2020.9074184.

[7] Mayowa, D.O. and Olajide, I.A., 2024. Design of a Voice Recognition System Using Artificial Neural Network. International Journal of Electrical and Computer Engineering Research, 4(1), pp.1-7..https://doi.org/10.53375/ijecer.2024.371

[8] Kumar, L., 2020. Desktop voice assistant using natural language processing (NLP). International Journal for Modern Trends in Science and Technology, 6(12), pp.332-335.https://doi.org/10.46501/ijmtst061262

[9] Akash, S., Jayaram, N. and Jesudoss, A., 2022, May. Desktop based smart voice assistant using python language integrated with arduino. In 2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 374-379). IEEE., Doi: https://10.1109/ICICCS53718.2022.9788267.

[10] Vora, J., Yadav, D., Jain, R. and Gupta, J., 2021. JARVIS: A PC Voice Assistant. International Journal of Advance Study and Research Work, 4(7), pp.386-393. Doi: https://doi.org/10.5281/zenodo.5323068

[11] Patil, A., Samant, S., Ramtekkar, M., Ragaji, S. and Khanapuri, J., 2020, April. Intelligent voice assistant. In Proceedings of the 3rd international conference on advances in science & technology (ICAST). Doi: http://dx.doi.org/10.2139/ssrn.3568721.

[**12**] Belinda, M.J., Nandhagopal, R. and Mahalakshmi, N.R., 2018. A voice-based text mail system for visually impaired. International Journal of Engineering and Technology (UAE), 7, pp.132-136. Doi: https://doi.org/10.14419/ijet.v7i1.7.10633

[13] Dias, P.M. and Jayakody, K., 2020, December. Virtual assistant in native language. In 2020 IEEE Asia-Pacific Conference on Geoscience, Electronics and Remote Sensing Technology (AGERS) (pp. 16-18). IEEE., Doi: https://10.1109/AGERS51788.2020.9452751.

[14] Singh, P., Nayak, P., Datta, A., Sani, D., Raghav, G. and Tejpal, R., 2019, February. Voice control device using raspberry pi. In 2019 Amity International Conference on Artificial Intelligence (AICAI) (pp. 723-728). IEEE. Doi: https://10.1109/AICAI.2019.8701409.

[15] Aktar, N., Jaharr, I. and Lala, B., 2019, February. Voice recognition based intelligent wheelchair and GPS tracking system. In 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE) (pp. 1-6). IEEE., Doi: https://10.1109/ECACE.2019.8679163.