International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

"JARVIS" - AI Voice Assistant

Rajat Sharma[1], Adweteeya Dwivedi[2]

[1]B-Tech, Department of CSE, SRM Institute of Science &Technology, Delhi NCR Campus, U.P., India
Email: *rrajat.sharma19[at]gmail.com*

[2]B-Tech, Department of CSE, SRM Institute of Science & Technology, Delhi NCR Campus, U.P., India
Email: *adweteeya1999[at]gmail.com*

Abstract: ***JARVIS, a virtual embedded voice assistant that includes cutting-edge technology based on gTTS and Python in developing a personalized assistant. JARVIS integrates the functionality of AIML and, together with Google, the industry leader, a text-to-speech platform and thus male/female voices into the gTTS libraries powered by the Marvel world. This is often the result of adopting the dynamic base Pyttsx Pythons considered wise in contiguous phases of gTTS, facilitating the establishment of essentially fine-tuned dialogues between assistants management and users. It will help end users in their daily activities like general human speech, query search in Google, Bing or Yahoo, video search, image retrieval, live weather, word meaning, predict and remind users of scheduled events and tasks. This is often the sole result of over-contributing by multiple contributors, such as AIML's usability and ability to dynamically merge with platforms like Python [pyttsx] and gTTS [Google Text to Speech]] results in the same JARVIS standard structure showing general reusability and almost zero or no maintainability.***

Keywords: Voice Assistant, NLP, Neural Network, Google Search

1. Introduction

AI voice assistant, also known as a virtual or digital assistant, is a device that uses voice recognition technology, natural language processing, and Artificial Intelligence (AI) to respond to people. Through technology, the device aggregates user messages, breaks them down, rates them, and gives meaningful feedback in return. Artificial intelligence can bring real conversations. Virtual assistants, understand natural language voice commands and performs tasks for users. These tasks, previously performed by a personal assistant or secretary, include dictation, reading text messages or exchanging email messages aloud, schedule appointments for end users. The AI assistant can also perform other activities, such as sending messages, answering phone calls, and getting directions. It also helps to read news and weather updates, open Google, You Tube, Stack Overflow, etc. , answer any questions, web scraping, play mu-sic, etc. Although this definition emphasizes the digital style of a virtual assistant, the term virtual assistant or virtual personal assistant is additionally unremarkably wont to describe contract employees United Nations agency work from home and perform body tasks un-remarkably performed by executives, assistant or secretary. Digital assistants can also be compared with other form of consumer-facing AI programming known as responsive advisors. Sensible adviser programs are topic-oriented, whereas virtual assistants are task-oriented.

"Virtual assistants are typically cloud-based programs that require internet-connected devices and/or applications to function". The technologies that power virtual assistants require vast amounts of knowledge, powering the platforms, as well as machine learning, language communication processes, and speech recognition arena. There are dedicated devices to provide virtual assistance. The most stylish on the market from Amazon, Google and Microsoft having Alexa, Google Siri and Cortana as AI voice assistants respectively given by each company.

AI voice assistants often perform simple tasks for end users, such as adding tasks to the calendar; provide information that can usually be searched in an Internet browser; or control and check the health of sensitive devices in the home, send emails, setting up of alarms, getting weather reports, can give your location, perform some basic mathematical calculations, check news, start the music, and open different websites like stack overflow, you tube, Facebook etc.

2. Related Work

2.1 Generalization

The below mentioned pie chart shows the analysis of virtual assistants in context to education as well as purpose of this work with a total of papers from 13 countries. The highest contribution was made by country England with most number of papers (3), followed by Russia and Switzerland (2 papers each). The remaining countries, namely Singapore, Pakistan, Canada, India, France, Bulgaria, Saudi Arabia and Germany are also mentioned with 1 paper each (Figure 2.1)

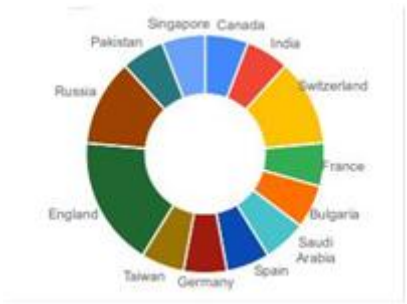


Figure 2.1: Pie Chart

The below displayed bar graph shows that growth is continuous in research papers since 2000, except the year 2010 (Figure 2.2). So this is indicating that this field of research is progressive in a contiguous manner.

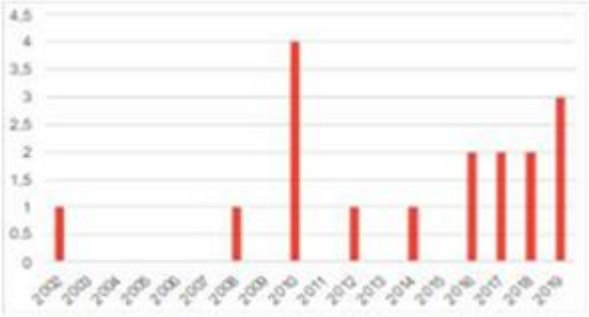


Figure 2.2: Bar Graph

2.2 Specific Researches

- AI technologies appear to be extensively adopted, folks don't use them in some cases. Technology adoption has been studied for several years, and there is a square measure, several general models, within the literature describing it. How-ever, having a lot of made-to-order models for rising technologies upon their options appears necessary. during this study, we have a tendency to develop an abstract model involving a replacement system quality construct, i.e., interaction quality, that we have a tendency to believe will higher describe the adoption of AI-based technologies.[5]
- Artificial Intelligence programs have currently become capable of difficult humans by providing professional Systems, Neural Networks (NN), Natural Language Processing (NLP), and Speech Recognition. Computer science brings a bright future for various technical inventions in various fields. This review paper shows the final thought of computer science, and therefore the use of speech recognition, and gifts the impact of computer science within the present and future world.[4]
- The project aims to develop a non-public assistant for Computers (computer Personal Assistant). It provides an easy interface for finishing a selection of tasks by using bound well-defined commands. Users will move with the assistant through voice commands. As a non-public assistant, it assists the end-user with regular activities sort as general human spoken communication, looking out queries, reading the most recent news, translating words, live weather, and causation mail through voice. The software package uses a device's electro-acoustic transducer to receive voice requests whereas the output takes place at the system's speaker.[11]
- "The virtual worlds offer many resources to engage their users (named avatars) like freedom of movement, teleport yourself to other locals, communication with other inhabitants (both text and voice messages), capacity to create, modify and destroy objects and the possibility of programming behaviors to these objects via scripts". The world is surplus of the resources for excelling in different fields but they just require some ways for communication. [10]
- This article introduces virtual embedded voice assistants including gTTS, and ad-vanced Python-based technology in custom assistant development. It integrates features of AIML and Google's industry-leading platform for text-to-speech con-version, and thus human voices are included in the gTTS library. This is often the result of applying the Pythons pyttsx dynamic base that is considered wise in the contiguous phases of gTTS and AIML, facilitating the establishment of noisy dialogues that are worth attention between the assistant and thus the user.[7]

The below survey table shows some projects with their respective pros and cons. (Table2.1)

Table 2.1: Survey Table

S. No	Project	Technologies	Result	Issues
1.	Voice Assistant using python	Voice activation, automatic speech recognition, dialog management	Design and implementation of digital assistance	Absence of additional or multiple features
2.	AI based voice assistant	Python 2.7 , Spider, json, machine learning	A modern model with some advance features established.	Similar with basic prototype and lacks multidimensionality
3.	An interpretation of AIML with integration of gTTS and Python	gTTS(Google text to speech), AIML(Artificial Intelligence Markup Language)	Integration of gTTS, AIML	Dependency on a particular platform
4.	Interoperability in virtual world	WWW(World wide web) services, HTTP, XML	Virtual world's communication, real world to virtual world (R2V)	Less vulnerable to modern operating systems
5.	Natural language understanding	Artificial Intelligence, Natural language processing	Understanding of natural language processing, syntact processing	Only developing the understanding of NLP, difficult to implement
6.	Chabot song recommender system	Python, chatterbot library, list trainer	Developing basic Chabot system	Dedicated to a particular feature only
7.	AI Chabot in python	Pip , NumPy, tensorflow, random	Automated communication system developed	Limited to certain queries and conversation

3. System Analysis

3.1 Training Model

- With the help of NN as neural network and NLP as natural language processing, create a brain of the model.
- And, with the help of machine learning modules and Deep Learning modules built emotions in the model and dataset to help the model in training.

3.2Neural Networks

"NN reflects the behavior of the human brain, enabling computer programs to recognize patterns and solve common

problems in artificial intelligence and other AI applications". An Artificial Neural Networks (ANNs) consists of a layer of nodes, including an input layer, one or more hidden layers, and an output layer. Each node is connected to another node, with weights and thresholds associated with it. If the output of an individual node is greater than the specified threshold, that node wakes up and sends data to the next layer of the network. Otherwise, the data will not be sent to the next layer of the network. The network relies on training datasets to learn and improve accuracy over time. (Figure 3.1)

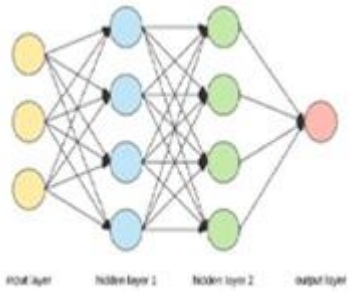


Figure 3.1: Neural Networks

Think of each node as a unique linear regression model consisting of input data, weights, bias as thresholds, and outputs.

ziri + th = z1r1 + z2r2 + z3r3 + th
output = g(r) = 1 if z1x1 + c >= 0; 0
if z1x1 + c < 0

When an input layer is specified, weight area units are assigned. These weights make it easy to see the importance of a particular variable. Large variables pay a lot of attention to the output for different inputs. Then the units of all input areas are incremented and summed with different weights. Then the output is passed. When this output exceeds a certain threshold, the node is triggered and knowledge is propagated to future layers in the network. This makes the exit of one node the entrance of future nodes. This method of passing knowledge from one layer to the future layer defines this neural network as a feed forward network.

3.4 Natural Language Processing

NLP implies "Natural Language Processing", which is part of the user language of computer science and one of the applications of artificial intelligence. This is a technology used by machines to understand, analyze, manipulate, and interpret human language. This helpdevelopers organize their knowledge to perform tasks such as translation, book reading, speech recognition, and topic segmentation. (Figure 3.2)

1) NLP helps users ask questions about a topic and get a direct answer within seconds.
2) NLP provides accurate answers to your questions. That is, it does not pro-vide unnecessary and unnecessary information.
3) NLP helps computers communicate with people in that language.
4) Most IT industries use natural language processing to improve the efficiency and accuracy of the documentation process and identify information from large databases.

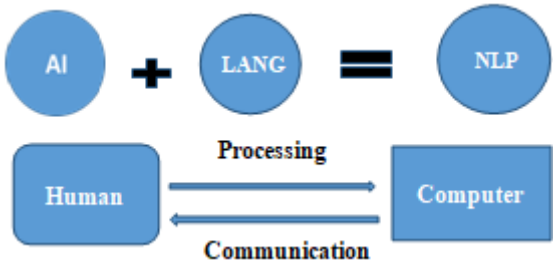


Figure 3.2: Natural Language Processing

3.3 Speech Recognition System

The speech recognition system is the core of the voice application system, which is capable of understanding the voice input given by the user, and at the same time operating the applications efficiently and generating voice feedback to the user. This system is an important component for users as a gateway to use their voice as an input component. (Figure 3.3) . In a word, in order to clearly recognize the user's speech command and get a response from the system, we should consider that the speech recognition system contains the whole process by which the application system directs the generation. voice signal to text data and some important meanings, forms of speech.

Figure 3.3: Speech Recognition System

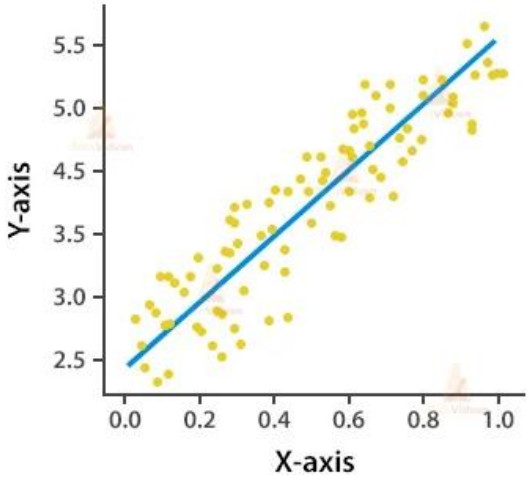
3.4 PyTorch - Machine Learning Library

PyTorch allows developers to teach neural network models in a very distributed way. It uses Python's native support for user-machine communication and asynchronous execution of aggregate operations to provide optimized performance in analytics and production environments.

- *torch.cuda: supports CUDA tensor types that implement the same function as CPU tensors.
- *torch.nn: this package provides many more classes and modules to implement and train the neural network.
- *torch.utilis.data: this package is mainly used for creating datasets.

3.5 Linear Regression Concept

This algorithm is a method of finding a linear relationship between a dependent variable and an independent variable by minimizing the distance. This is a supervised algorithm. Here, we use a machine learning supervised algorithmic approach to categorize individual categories. Using this algorithm, we created a voice assistant model that allows users to predict relationships between dependent and independent entities.



$$D=p+qI$$
$$p= (D)(I^2)-(I)(ID)n(I^2) - (I)^2(D)(I^2) - (I)(ID)n(I^2) - (I)^2$$
$$q= n(ID) - (I)(D)n(I^2) - (I)^2$$

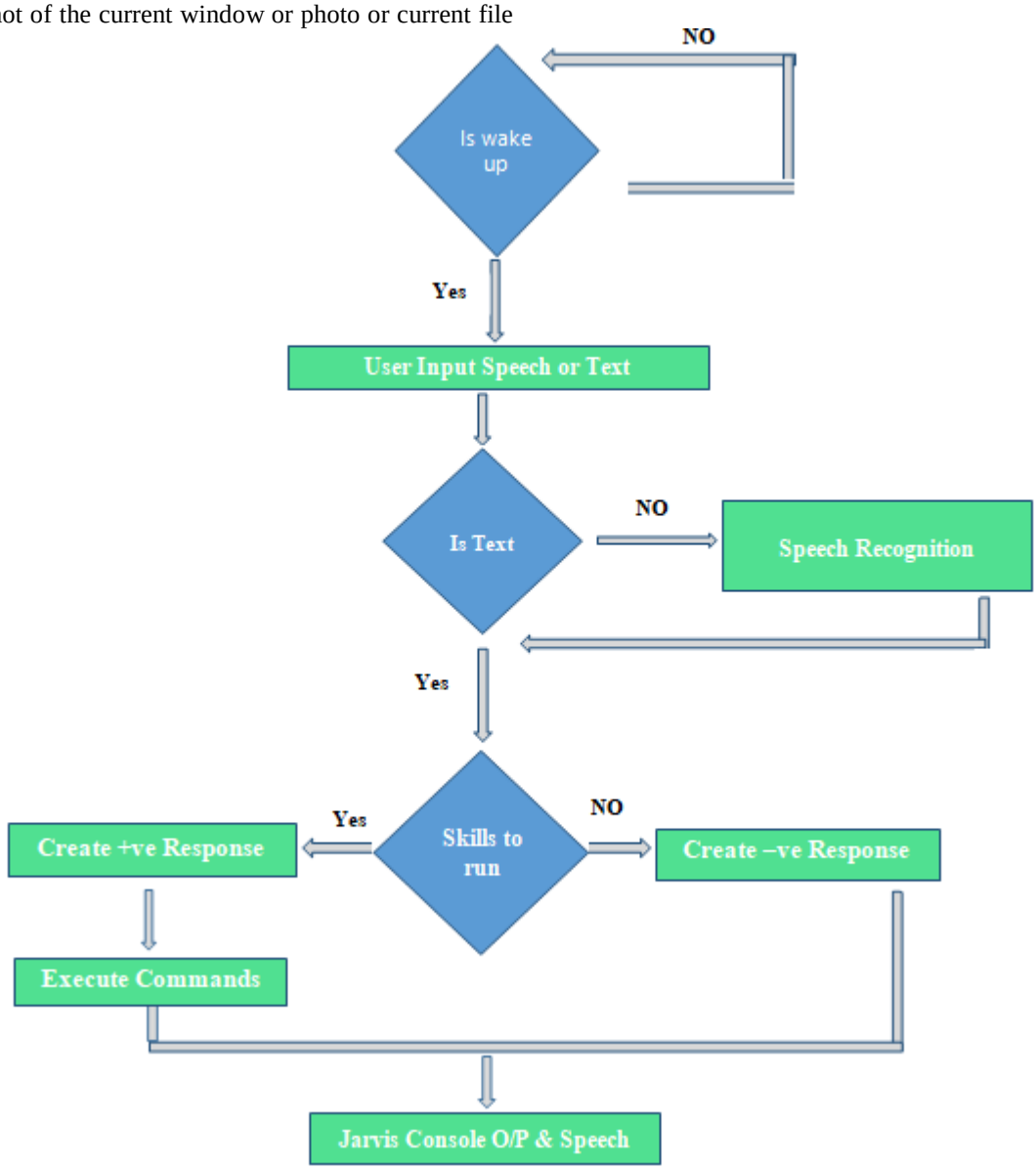
I is the independent variable
D is the dependent variable
p is intercept and q is slope of line here

4. Proposed System

- The voice assistant initiates voice mode and prompts the user to provide input in voice/text format for best results from the voice assistant. As this program can also be controlled with your phone with help of an application '**WO-MIC'**, it just turns any android phone into a wireless microphone and helps in the reduction of unwanted noise in the environment.
- Using this application, which is **Wikipedia's search engine**, users can contact the wizard and the wizard will retrieve the data from the internet. The results are displayed in the console window in audible format, up to a limited number of lines.
- **Getting Current News** about his/her motherland, about world, about technologies, about sports or about entertainment of the industry and much more, the user can easily get the news just by giving voice input to assistant to open news so it will open new tab and it can also fetch the data from the websites and return it to the console and read out for user without any labor.
- **Weather Forecast**, through this feature users can see the weather forecast for any location. In addition, the temperature and humidity of Kelvin will return the weather.
- **Open Applications** like , YouTube, google search engine , launching websites , system applications with the help of web browser python library and os for opening system applications(like, code editor, notepad,chrome,etc.)
- **Close Applications**, the application work perfectly by providing a command 'TASKKILL/ F/im file.exe'. The assistant close that application asked to close.
- **Automation**, the application performs automation for YouTube and any search engine with the help of keyboard python library. The user just need to give input and the assistant will perform the automation ask.
- **Voice Assistant caneven repeat the user's words** by takeCommand function and speak function.
- **WhatsApp Messages**, the application work by taking mobile number of the receiver or the name of the receiver, message to send , time when to send as a query. As the result , voice assistant will send the message and inform you. This is done with help of pywhatkit python library. And the history of messages will be saved in pywhatkit database file .
- **Checking Internet Speed**, the application is done with the help of speedtest python library by which assistant will check and return the result on the console.
- **Checking my location**, this feature allows users to view their current location or find directions to any location.
- **Listening to music**, the voice assistant plays the music requested by the user, either from the user's system or through an online search, without the user having to do it.
- Jarvis translator, this feature translates the user's original text input into the desired language. Over forty human languages are stored in the dictionary.
- **Audiobooks**, the application is very attractive as the voice assistant will open and read the book in your favourite language for the user to understand the book with the help of pypdf python library.
- The **Assistant create a note** to save the user's important data for future use.
- **Sending Mails**, this feature allows users to send an email to someone whose contacts include an email address. It then sends the successful execution of the task back to the user via the Hearing Assistant.
- **TimeTable Notification**, the voice assistant will remind you the work accord-ing to the user's time table schedule and as a result it also give notification with the help of notify python library .
- The **Voice Assistant can answer any query** with the help of wikihow python library and wolframalpha algorithm.

- A **setting alarm** is a basic function of any device, this allows the user to set the alarm to a specific time.
- **Chatbot**, this feature communicates with the user on a case-by-case basis. It also works whenever the user provides voice input to the assistant and the user receives the output in the voice response of the voice assistant ChatBot.
- **ScreenShot**, this feature allows the user to take a screenshot of the current window or photo or current file and ask the user for the name of the file to save in the required file folder on the system for later viewing.
- **Calculations**, this function performs an arithmetic calculation with a user's voice command and produces an output that is a calculated solution through a voice assistant.

4.1 System Architecture



Initially the condition is that if the Jarvis voice assistant is active or not, if it is active then it asks for the user input otherwise make jarvis active(make it on). Then user provides the input in the form of speech or text, after that if the input provided is in text then it goes for the action to be taken or the skills to be executed, else if the input is in speech then it uses the speech recoginition feature and converts it into text and goes for the action. (Figure 4.1)

Now after the proceedings if the skills to be executed are adequate to Jarvis then it gives a positive response to the user in form of speech and then executes the commands for operations, hence gives the console output and speech.On the other hand if the skills to be executed are not adequate or inappropriate to Jarvis it gives a negative response and executes no further commands to give console output. (Figure 4.1)

4.2Sequence Diagram

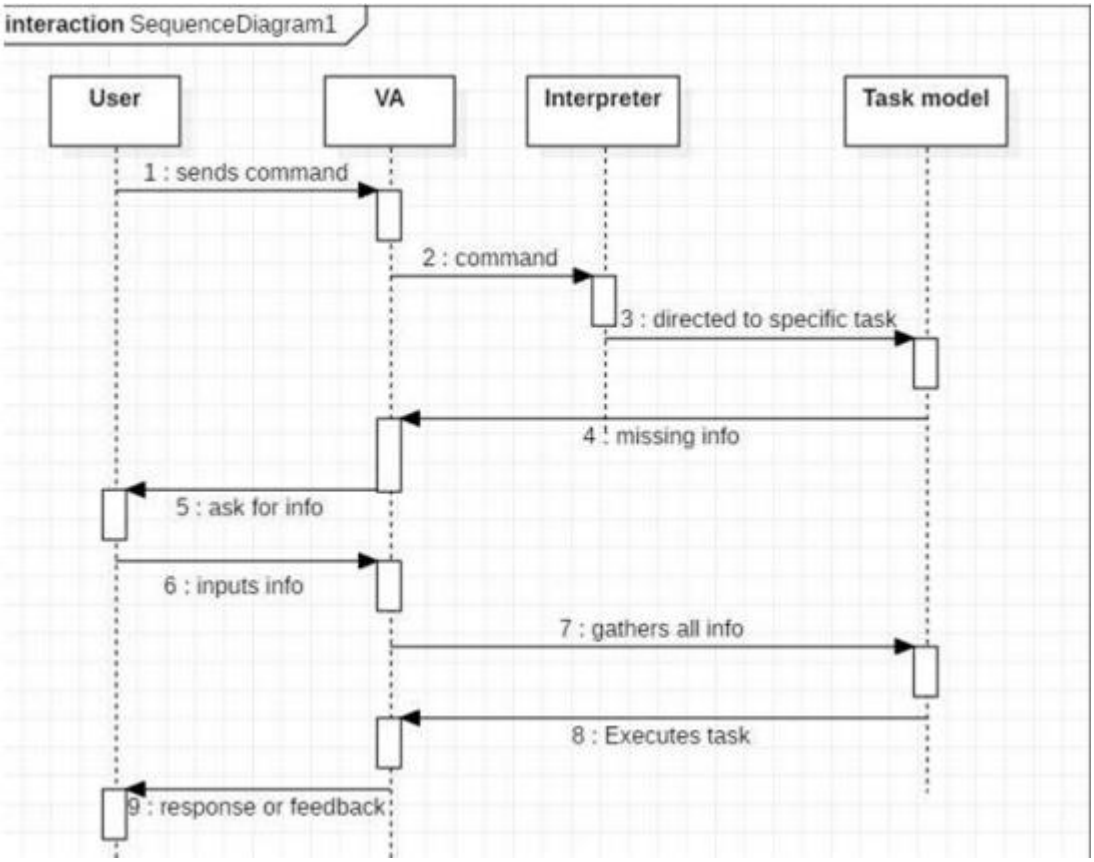


Figure 4.2: Sequence Diagram

- The user sends command to the voice assistant Jarvis then it forwards it to Interpreter i.e. speech recoginition feature here and then is directed perform the specific task, after the processing in task model Jarvis executes the task and give the response or feedback to the user. (Figure 4.2)
- If after the processing at task model there is some missing information then Jarvis asks for that information, takes the input again, gathers all information and follow the same process as detailed above. (Figure 4.2)

4.3Use Case Diagram

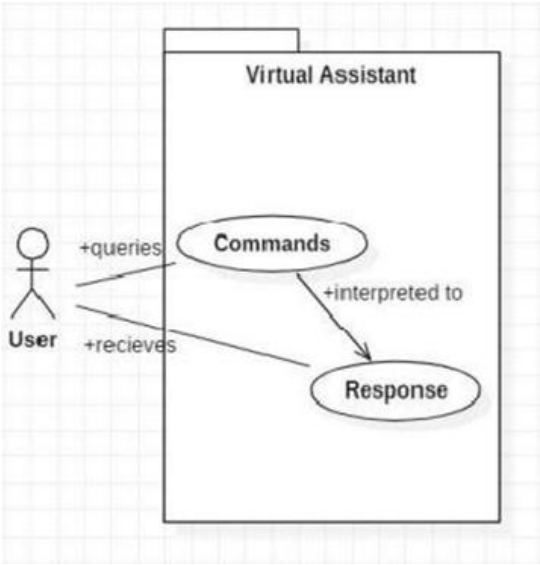


Figure 4.3: Use Case Diagram

4.4 Activity Diagram

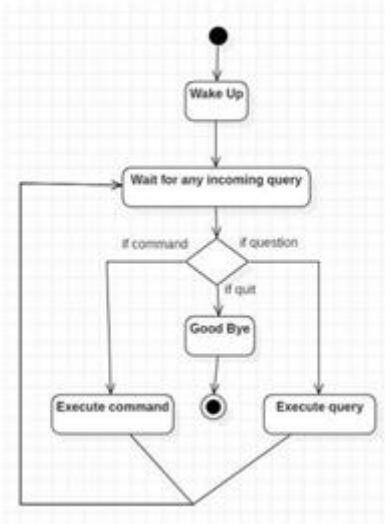


Figure 4.4: Activity Diagram

5. Experimental Result

On User speech command voice assistant display google search of the query asked and read the solution for the user too.(Figure 5.1)

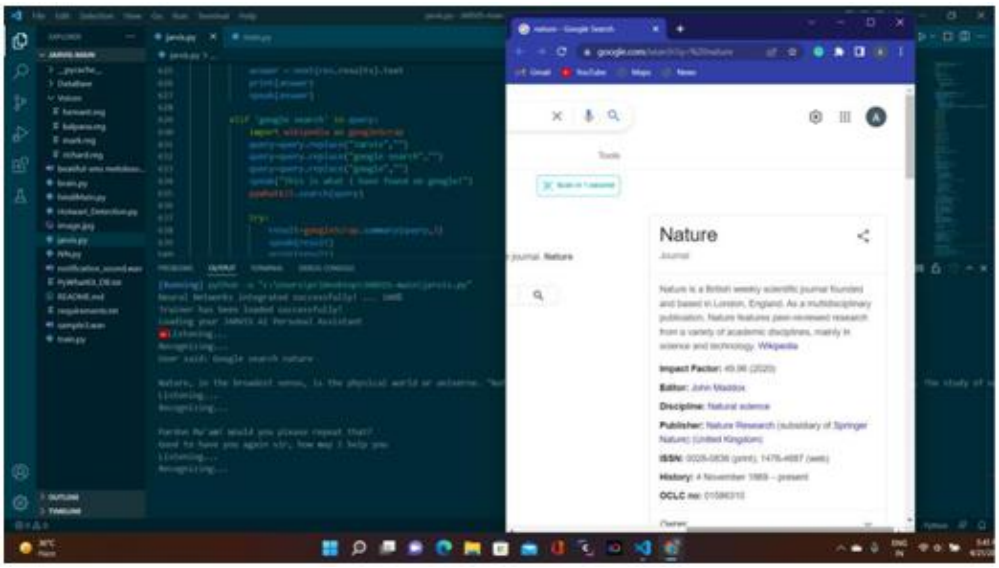


Figure 5.1: Google Search

On User speech command voice assistant open News (Figure 5.2)

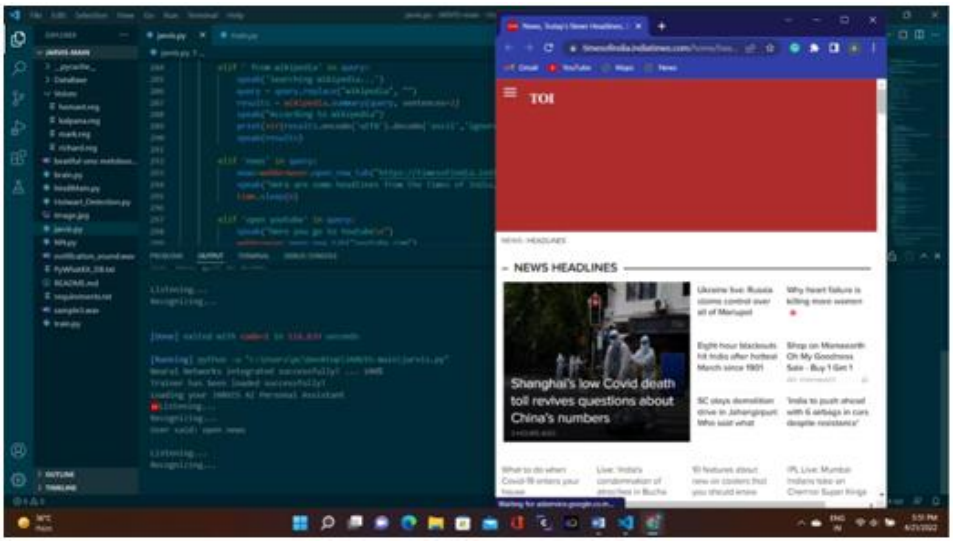


Figure 5.2: Open News (TOI)

On User speech command voice assistant open 'my location' (Figure 5.3)

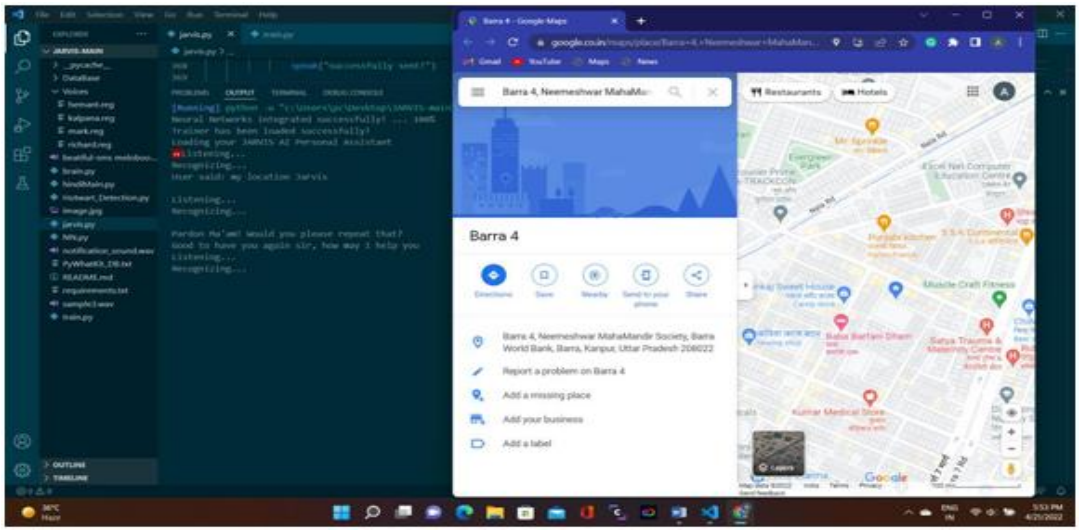


Figure 5.3: Open My Location

6. Conclusion

Jarvis - An AI Voice Assistant System uses speech recoginition, gTTs and other AI techniques along with Neural Networks and Natural Language Processing for a smart responsive system to the given circumstances or conditions. It can reduce the workload of basic human activities or the daily activities and can replace some human working posts like personal secretaries employed for scheduling a person's per day time table. Critically, the system is designed to interrelate with other sub-systems smartly and comprehensively.

The system will have the following phases: Input phase in which data or query given in form of text or speech, Interpretetion of voice to text, Processing and storing of data, producing output in the form of voice from the refined text to Jarvis console. The information produced at each step can then be used to retreive patterns and analyze them for later use. This could be the main basis for artificial intelligence machines to learn and recognize patterns for people. So, based on a literature study and analysis of persisting systems, the conclusion is derived that our provided system will not only facilitate interaction with systems and modules but also keep users more organized.

7. Future Enhancement

Enhancement in the capacity of database or the data training sets can be done in this for more situations or the acquaintances that can be faced by JARVIS. This would upgrade its effectiveness and the wide range ability of producing responses. Further addition of more voices can also be done as an additional feature. So these limitations can be broken with the increase in data training sets.

The interface of the system can be improved more or we can say can be optimized. From saying more optimized it is meant that the interface can be more user friendly, comprehensive and easy to use for more percentage of users. So the Jarvis would become more accessible and intractable.

References

[1] Alotto, F., Scidà, I., and Osello, A. (2020). "Building modeling with artificial intelligence and speech recognition for learning purpose." Proceedings of EDULEARN20 Conference, Vol. 6. 7th.
[2] Beirl, D., Rogers, Y., and Yuill, N. (2019). "Using voice assistant skills in family life." Computer-Supported Collaborative Learning Conference, CSCL, Vol. 1, Inter-national Society of the Learning Sciences, Inc. 96–103.
[3] Canbek, N. G. and Mutlu, M. E. (2016). "On the track of artificial intelligence: Learning with intelligent personal assistants." Journal of Human Sciences, 13(1), 592–601.
[4] Malodia, S., Islam, N., Kaur, P., and Dhir, A. (2021). "Why do people use artificial intelligence (AI)-enabled voice assistants?." IEEE Transactions on Engineering Management.
[5] Nasirian, F., Ahmadian, M., and Lee, O.-K. D. (2017). "Ai-based voice assistant systems: evaluating from the interaction and trust perspectives.
[6] RAJA, K. D. P. R. A. (2020). "Jarvis ai using python.
[7] Sangpal, R., Gawand, T., Vaykar, S., and Madhavi, N. (2019). "Jarvis: An inter-pretation of AIML with integration of gtts and python." 2019 2nd International Con-ference on Intelligent Computing, Instrumentation and Control Technologies (ICI-CICT), Vol. 1. 486–489.
[8] Steen, J. and Wilroth, M. (2021). "Adaptive voice control system using ai.
[9] Terzopoulos, G. and Satratzemi, M. (2019). "Voice assistants and artificial intelligence in education." Proceedings of the 9th Balkan Conference on Informatics. 1–6.
[10] Tibola, L. R. and Tarouco, L. M. R. (2013). "Interoperability in virtual world." XVIII Congreso Argentino de Ciencias de la Computación.
[11] Vora, J., Yadav, D., Jain, R., and Gupta, J. (2021). "Jarvis: A pc voice assistant.
[12] Nasirian et al. (2017) Malodia et al. (2021) Vora et al. (2021) Tibola and Tarouco(2013) Sangpal et al. (2019) RAJA (2020) Beirl et al. (2019) Terzopoulos and Satratzemi
[13] (2019) Alotto et al. (2020) Steen and Wilroth (2021) Canbek and Mutlu (2016)