Harvey- The Python-Based Intelligent Voice Assistant

Shivansh Bhat[1], Shourya Goswami[2], Rahil Bopche[3], Veena Jadhav[4], Rohini Jadhav[5]

[123]*Student, Computer Engineering Department, BVDUCOE Pune-43, Maharashtra, India*

[4]*Assistant Professor, Bharati Vidyapeeth University College of Engineering Pune*

[5] *Associate Professor, Bharati Vidyapeeth University College of Engineering Pune*

Abstract - **Today, As the world is making advancements in the areas of technology, various tasks can be accomplished using machines that were once impossible to do. One such major advancement made is bringing in automation in day-to-day devices to reduce human efforts in performing everyday tasks, using smart devices. One of the most popular real-time applications of automation is the virtual assistant which allows the user to perform tasks by the means of voice recognition. This Survey paper will define the working and features of the voice assistant- Harvey, which will be majorly used for triggering windows native features as well as to overcome the limitations of pre-existing voice assistants for windows.**

1.INTRODUCTION

In this new era of expeditious technology, the industry is gradually shifting towards automation by adopting smart virtual assistants in digital devices like smartphones, laptops, tablets, etc. These virtual assistants facilitate various day-to-day tasks pertaining to a smart device, by making use of voice recognition and speech identification, with minimal human efforts, in order to make the user experience of these devices and their applications more efficient and convenient to use.

Currently, virtual assistants are very popular among all categories of users as they've become nearly successful in being personal companions to the users by making their lives easier and saving their efforts to operate a computerized device. Using this technology, users can control the functionalities of a device, surf the internet with layman knowledge about the technology and even interact or ask questions from the machines regarding any relevant topic.

Some popular pre-existing virtual assistants, currently in use include Apple's Siri, Android's Google Assistant, and Amazon service's Alexa. These technologies have highly attracted the population to switch to smart devices and benefit from the features they have to offer.

Harvey- the desktop voice assistant specially designed to perform windows native operations using voice recognition mechanism successfully offers the following features: -

1 Recognizes different voices
2 Enables the user to mute, unmute, volume up and down, play and pause, backward, and forward any audio media.
3 Sends across Emails
4 Shows the weather forecast across the globe
5 Recites jokes
6 Opens websites
7 Communicates with the user
8 Searches for a particular topic as per user's choice, on Wikipedia

2.LITERATURE SURVEY

Voice assistants are rapidly gaining popularity among the masses as a virtue of their demand in certain electronic devices like Bluetooth headsets, speakers, music systems, desktops, laptops, etc. Nowadays, an enormous amount of data is generated, hence for the voice assistant to handle such large chunks of data, the concepts of machine learning are incorporated in them to train them accordingly. To make the most out of their performance, voice assistants are also exposed to technologies other than ML such as IoT, NLP, and big data access management. In the paper [1] "Speech recognition using flat models" by Patrick Nguyen and all, a novel direct modeling approach for speech recognition is being put forward which helps to ease out the measure of consistency in the sentence spoken by the user. They have entitled this approach as Flat Direct Model (FDM).[2]

The application of voice assistants in today's market has gone up to a much higher level in the paper [3] "Smart Home Using Internet of Things" by Keerthana S in which they've discussed how making use of smart assistants plays a role in developing a smart home system using Wi-Fi network and IoT which enables controlling all the features of the home by the means of voice commands. The in-depth real-time applications of voice assistants have been discussed in the paper [4] "An Intelligent Voice Assistant Using Android Platform" by Sutar Shekhar, which majorly stresses the fact that mobile users, with the help of voice commands can perform tasks without having to type or make any effort. The association of Natural Language Processing (NLP) in voice assistants is immensely required which may lead to the creation of a trendsetting assistant.

This area of the new age of virtual assistants with power speech recognition has seen some major improvements and discoveries. This is to a large extent due to its demand for new-age devices like new smartwatches or high-tech fitness bands, speakers, laptops or desktops, Bluetooth earphones, smartphones, television, etc.

Almost every new-age digital device is getting imported with voice assistant technology which helps users to control using only speech recognition. Every day, new technology is being developed to improve the experience of users and the performance of automated voice search.

[5]   original ref- REF- J. B. Allen, "From lord Rayleigh to Shannon: How do humans decode speech," in International Conference on Acoustics, Speech and Signal Processing, 2002.

Language is the most important medium of communication, according to J. B. Allen et al, and speech is its primary interface. In the human-computer interface, the spoken signal was translated into analog by the equipment and machine's understanding of digital waveform. A widely used technology with endless applications. Speech technologies allow machines to communicate and react appositely to humans in a consistent and acceptable manner, also essential and appreciated services are provided. A synopsis of the speech is included in the research process for identification, its core model, and its working approaches, and a reasonable description of research into a variety of approaches used for systems

for voice recognition. SRS is improving gradually and has an unlimited number of uses.

[6]   - Ref 10 - Tanja Schultz and Alex Waibel, discussed concerns about the expansion of voice technology products around the world, the immovableness to new destination languages points out to be a significant concern. Importance, the emphasized research on the question of the way to import large vocabulary continuous speech recognition (LVCSR) systems in an exceedingly quick and well-organized way. More significantly the research must conclude acoustic models for a completely unique language using speech data from totally different source languages, but solely restricted information from the source language verification results from various acoustic models like language-adaptive, independent, and dependent are delineated and deliberated within the framework of Global Phone project that examines LVCSR methods in fifteen languages.

[7]   Cortana (existing voice assistant for the Windows operating system) uses a Microsoft translator back-end to perform this magical feat. Microsoft translator speech API (Application Program Interface) is a cloud-based automatic translation service. Microsoft translator's main and important mission is to break the language barrier by providing translation. It uses Bing as a web search engine which has been developed using ASP.NET. It was Tellme Networks, bought by Microsoft in 2007, from where the natural processing capabilities of Cortana are derived & are coupled with a semantic search database called Satori. Its ultimate motive is to get more meaning from data and integrate rational concepts with more powerful abstraction concepts that we know from the field of artificial intelligence. The aim is to provide high- level modeling primitives as an integral part of a data model to facilitate the representation of real-world situations.

[8]   Cortana's performance is powered by a Microsoft translation backend. The primary goal of Microsoft translators is to break down linguistic barriers through translation. It makes use of the Bing web search engine, which was created with ASP.NET. In 2007, Microsoft purchased Tellme Networks, from which Cortana's natural processing skills are derived and paired with Satori, a semantic search database. Its ultimate goal is to

make data more meaningful by combining logical conceptions with more powerful abstraction concepts learned via AI. The goal is to give high-level modeling primitives as an inherent part of a data model so that real-world scenarios can be represented.

3.PROPOSED STRUCTURE
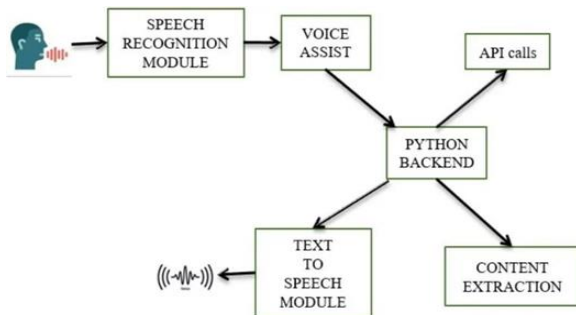
The current system makes use of Google's online speech recognition system to convert the input speech by the user to text. The input speech might be able to obtain texts from a particular corpus being organized on the PC network server with the help of a microphone and is temporarily stored in the system post which it is sent to google cloud for the purpose of speech recognition. Post this, the corresponding text is received and further given to the central processor. Then python backend is responsible to get the output from the module of speech recognition which is then identified on the grounds of whether the type of command is an API call or context extraction. After this, the output is sent back to the backend in order to provide the desired output to the user. The API comes into use, and when simplified, refers to Application Programming Interface. It is basically a software intermediary which is responsible for allowing certain applications to interact with one another. In simpler words, it acts as a messenger that helps to deliver the user's request to the respective provider, whom the user is requesting, and then delivers the desired response back to the user. CE or Context extraction refers to the task of extracting information automatically in its structured form from unstructured/semi-structured documents in their machine-readable format. In frequent cases, this particular activity is concerned with proceeding with human-understandable texts using an NLP (Natural Language Processing).

The test results, in this case, would include the processing of multimedia documents out of images or video as context extraction. At last, TTS or Text-to-Speech is the capacity of PCs to read out the text. The written text is converted to a phonemic representation using a TTS engine which is responsible for it. This phonemic representation is then converted into waveforms that will be the output in the form of sound. These TTS engines are made available with different languages and broadened vocabularies through third-party publishers.

4. METHODOLOGIES

3.1 GTTS
Also known as Google Text-to-Speech. It is a library used in Python and interacts with Google Translate text-to- speech API. The GTTS library will be imported from the GTTS module, which may be used for voice translation. This module is used by Harvey to convert text to speech as it is programmed in text.

3.2 Speech recognition
The capacity of a machine to listen to spoken words and recognize them is known as speech recognition. The uttered words may then be converted to text, a query can be made or a response can be given using Python's voice recognition. This module bridges the gap between the user and the voice assistant.
With this module, Harvey recognizes the voice and converts it into words for it to understand and act accordingly.

3.3 OS
The Python OS module includes methods for creating and deleting directories (folders), retrieving their contents, altering and identifying the current directory, and more.

This module enables Harvey to interact with the underlying operating system.

3.4 R.E (Regular Expressions)

It's a particular character sequence that searches for a string or collection of strings using a search pattern. It can match a pattern to detect the presence or absence of text, and it can also divide a pattern into one or more sub-patterns.

3.5 Web browser

The web browser module gives you access to the system's default web browser. This module has an open function that opens a file or a URL, searches for it, and displays it in the browse.

This module enables Harvey to access any file or website via the default web browser and give results.

3.6 SMTPlib

The SMTPlib is a library in Python which is responsible for sending across emails by making use of the Simple Mail Transfer Protocol (SMTP).

The SMTPlib, being a built-in module, does not require installation. It helps to eliminate all the complexities of SMTP.

It enables Harvey to send an email to someone with just the user's voice commands.

3.7 Requests

The requests module enables the user to send across HTTP requests by making use of Python. A response object is returned by the HTTP request which contains all the response data related to the content, status, or encoding.
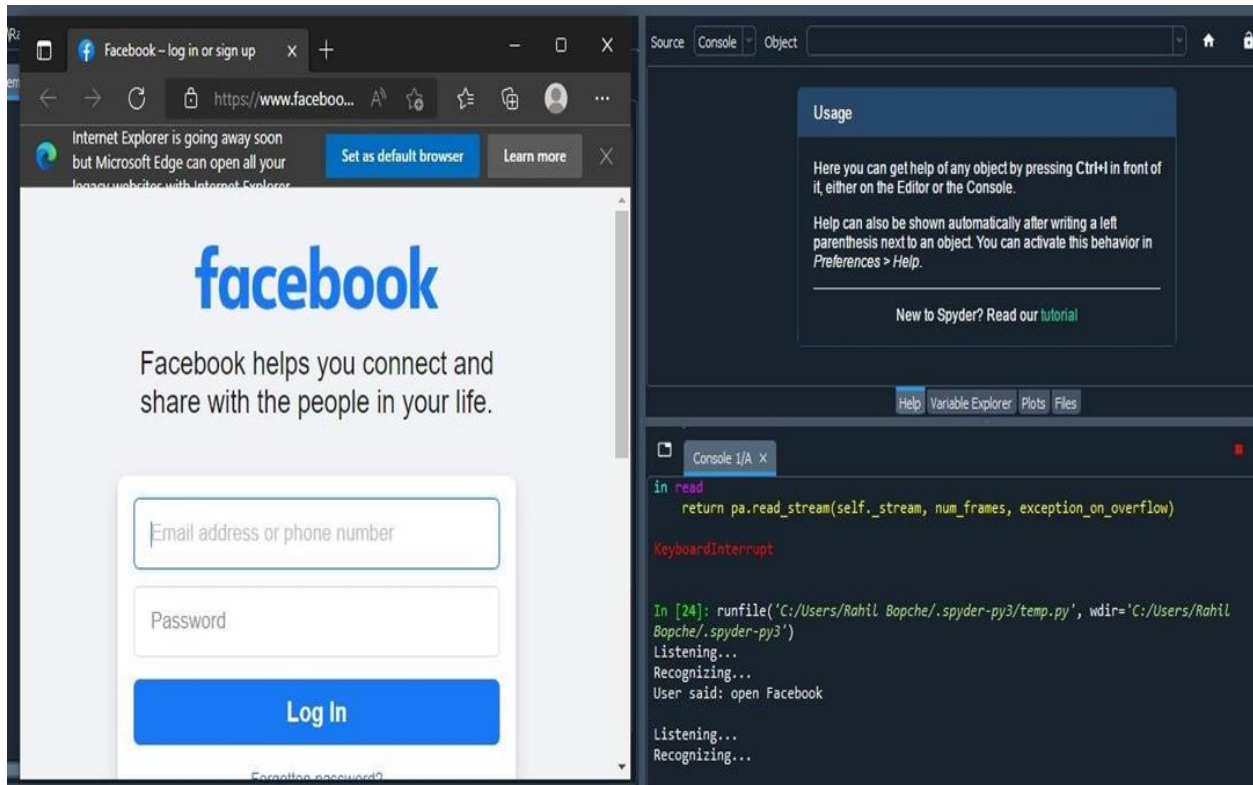
3.8 Weather API

The Weather API is an API which helps Harvey to get weather information from any place. It fetches the data using an API key which is provided on the website. The data is fetched by requesting that API key and then use data is extracted to give the results. This API gives real- time data at any point in time of any place.

5. RESULT
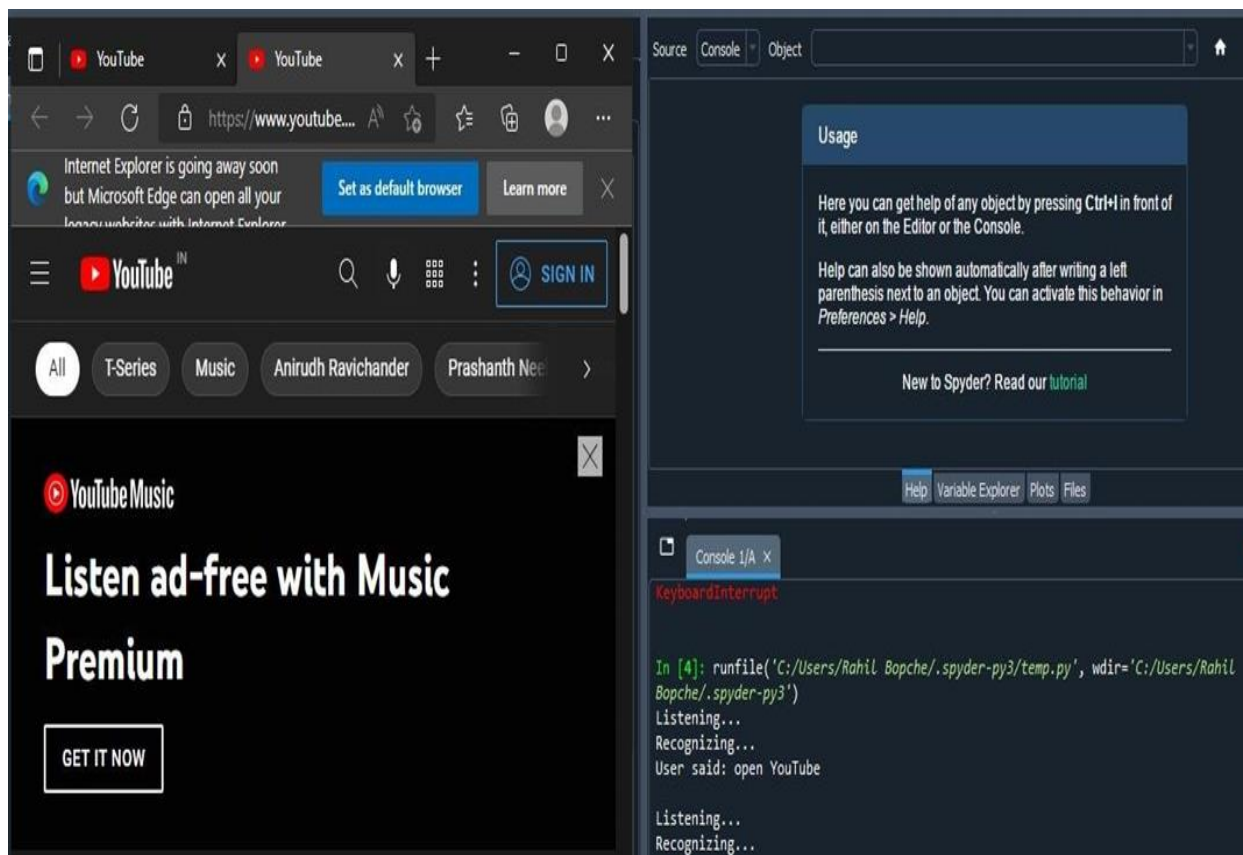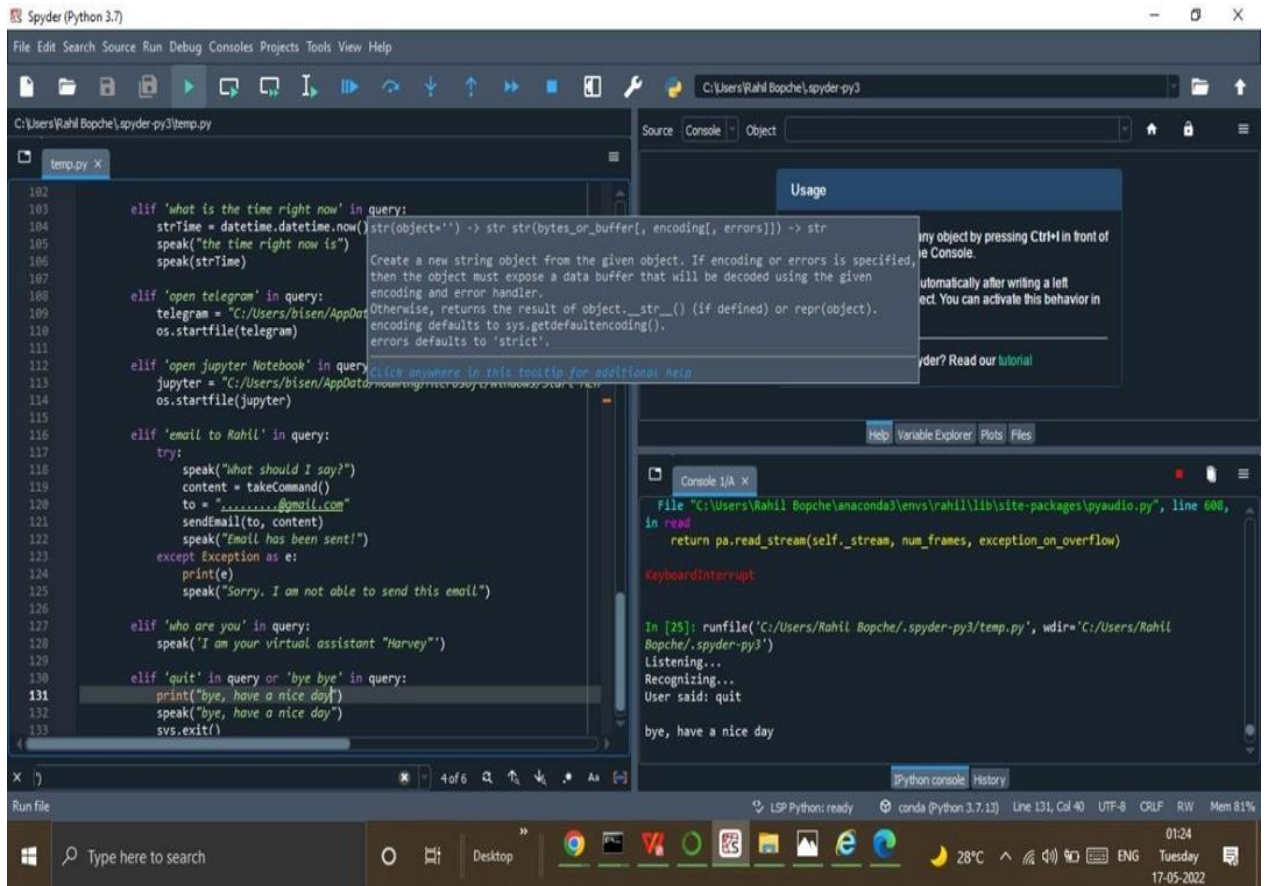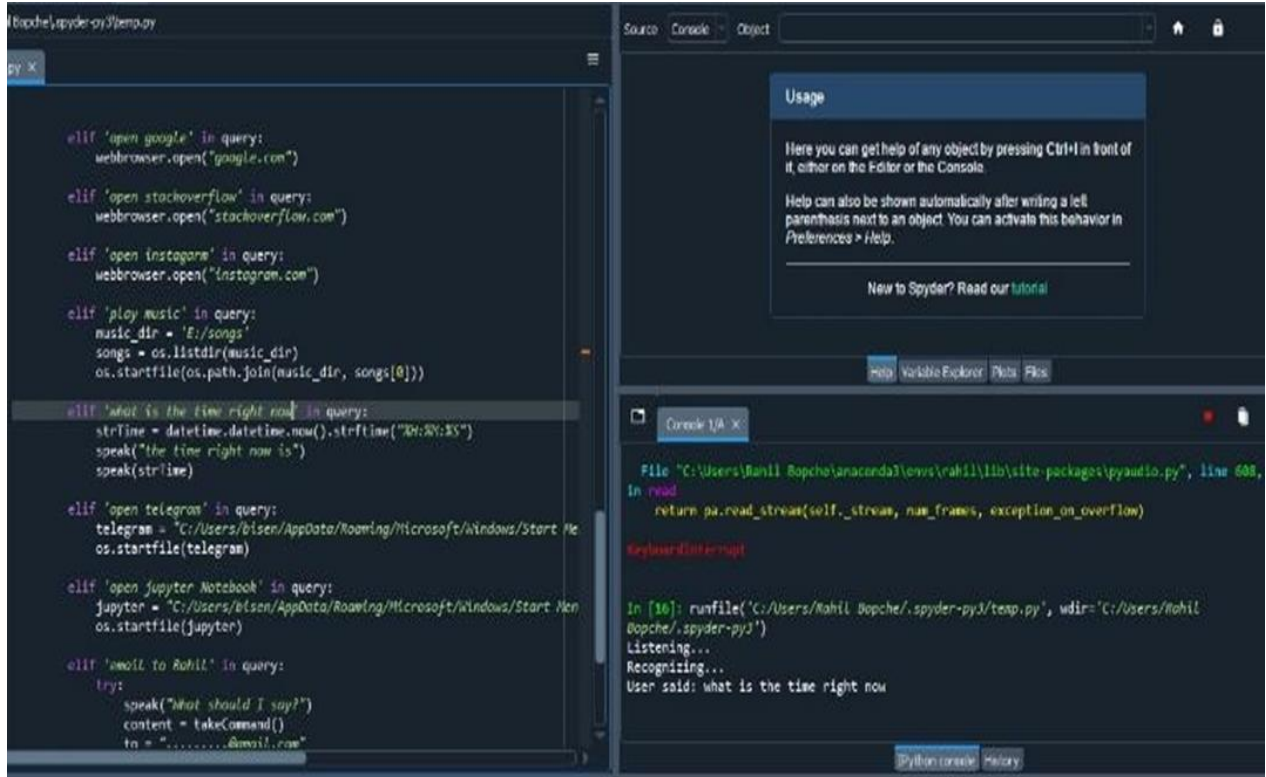
6.CONCLUSION

In this research paper, we have discussed in-depth Harvey- the desktop voice assistant developed using various modules in python. This assistant currently works with the help of the internet and is successful in performing numerous tasks including weather updates, opening various websites on the web browser as per the user's choice, sends email plays music, searches for different topics on Wikipedia, opens general desktop applications and facilitates general conversation with the user, and generates voice speech depending upon the user's response.

The current version of this assistant offers a good user experience and is quite responsive, though some fields require improvements in the overall functionalities and will be incorporated with future software updates.

7.FUTURE SCOPE

The future scope for our voice assistant according to our research would be: -

1   Harvey can be released as open-source software in a bunch of communities so that contributions can be made by developers worldwide.

2   Harvey could be later provided with AI & ML algorithms for a better understanding and efficient performance.

3   Harvey could be later integrated with the raspberry pi and could be innovated as a new device like Amazon Alexa and Google Home.

4   Harvey could get better at recognizing different voices for a better-personalized experience.

5   In the future, Harvey could be more protected and secured by recognizing specific voices and identifying unknown voices.

6   Harvey could also be designed with push notifications via connecting with third-party apps.
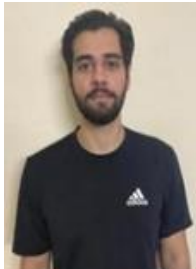
REFERENCES

[1]   "Speech recognition using flat models" by Patrick Nguyen and all.

[2]   A Novel Python-based Voice Assistance System for reducing the Hardware Dependency of Modern Age Physical Servers; International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395- 0056 Volume: 08 Issue: 05 | May 2021 www.irjet.net p-ISSN: 2395-0072

[3] "Smart Home Using Internet of Things" by Keerthana S.

[4] "An Intelligent Voice Assistant Using Android Platform" by Sutar Shekhar.

[5] REF- J. B. Allen, "From lord Rayleigh to Shannon: How do humans decode speech," in International Conference on Acoustics, Speech and Signal Processing, 2002.

[6] Voice Assistant Using Python; © July 2021| IJIRT | Volume 8 Issue 2 | ISSN: 2349-6002; Nivedita Singh, 2Dr. Diwakar Yagyasen, 3Mr. Surya Vikram Singh, 4Gaurav Kumar, 5Harshit Agrawal1,2,3,4,5, Department of CSE, Babu Banarasi Das National Institute of Technology and Management, Lucknow, India.

[7] CORTANA-INTELLIGENT PERSONAL DIGITAL ASSISTANT: A REVIEW Volume 8, No. 7, July – August 2017 International Journal of Advanced Research in Computer Science.

[8] Desktop Virtual Assistant; International Journal for Research in Applied Science & Engineering Technology; Paper ID: IJRASET407.

BIOGRAPHIES

Mr. Shivansh Bhat is pursuing Bachelor of Technology in Computer Engineering at the Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune. His research interest includes Machine Learning using python and Web development.

Mr. Rahil Bopche is pursuing Bachelor of Technology in Computer Engineering at the Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune. His research interest includes Machine Learning using python.

Mr. Shourya Goswami is pursuing Bachelor of Technology in Computer Engineering at the Bharati Vidyapeeth Deemed to be University) College of Engineering, Pune. His research interest includes Machine Learning using Python and Artificial Intelligence

IN COLLABORATION WITH PROFESSOR:

1. Veena Jadhav, Assistant Professor, Bharati Vidyapeeth (Deemed to be) University College of Engineering Pune.
2. Dr. Rohini Jadhav , Associate Professor, Bharati Vidyapeeth (Deemed to be) University College of Engineering Pune.