**VIRTUAL PERSONAL ASSISTANT USING RPA, AI AND PYTHON**

A Project-II Report

Submitted in partial fulfillment of requirement of the

Degree of

**BACHELOR OF TECHNOLOGY in COMPUTER SCIENCE & ENGINEERING**

BY

**Pranav Gautam EN19CS3T1002**

**Vivek Patidar EN18CS301291**

Under the Guidance of

**Mr. Prasanna Kapse**



**Department of Computer Science & Engineering**

**Faculty of Engineering**

**MEDI-CAPS UNIVERSITY, INDORE- 453331**

**April 2022**

**REPORT APPROVAL**

The project work **“Virtual Personal Assistant using RPA, AI and Python”** is hereby approved as a creditable study of an engineering subject carried out and presented in a manner satisfactory to warrant its acceptance as prerequisite for the Degree for which it has been submitted.

It is to be understood that by this approval the undersigned do not endorse or approve any statement made, opinion expressed, or conclusion drawn therein; but approve the “Project Report” only for the purpose for which it has been submitted.

Internal Examiner

Name:

Designation

Affiliation

External Examiner

Name:

Designation

Affiliation

**DECLARATION**

We hereby declare that the project entitled **“Virtual Personal Assistant using RPA, AI and Python”** submittedin partial fulfillment for the award of the degree of Bachelor of Technology in ‘Computer Science & Engineering’ completed under the supervision of **Mr. Prasanna Kapse, Assistant Professor, Computer Science & Engineering,** Faculty of Engineering, Medi-Caps University Indore is an authentic work.

Further, we declare that the content of this Project work, in full or in parts, have neither been taken from any other source nor have been submitted to any other Institute or University for the award of any degree or diploma.

Pranav Gautam EN19CS3T1002

Vivek Patidar EN18CS301291

**5 May 2022**

**CERTIFICATE**

I, **Prasanna Kapse** certify that the project entitled **“Virtual Personal Assistant using RPA, AI and Python”** submittedin partial fulfillment for the award of the degree of Bachelor of Technology in Computer Science & Engineering by **Pranav Gautam and Vivek Patidar** istherecordcarried out by them under my guidance and that the work has not formed the basis of award of any other degree elsewhere.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Prasanna Kapse

Assistant Professor

Computer Science & Engineering

Medi-Caps University, Indore

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Dr. Pramod S. Nair

Head of the Department

Computer Science & Engineering

Medi-Caps University, Indore

**ACKNOWLEDGEMENTS**

We would like to express my deepest gratitude to the Honorable Chancellor, **Shri R. C. Mittal,** who has provided us with every facility to successfully carry out this project, and our profound indebtedness to **Prof. (Dr.) Dilip. K. Patnayak,** Vice Chancellor, Medi-Caps University, whose unfailing support and enthusiasm has always boosted up our morale. I also thank **Prof. (Dr.) D. K. Panda,** Pro Vice Chancellor, **Dr. Suresh Jain,** DeanFaculty of Engineering, Medi-Caps University, for giving us a chance to work on this project. We would also like to thank our Head of the Department **Dr. Pramod S. Nair** for his continuous encouragement for betterment of the project.

I express my heartfelt gratitude to my Internal Guide **Mr. Prasanna Kapse** without whose continuous help and support, this project would ever have reached completion.

**Pranav Gautam**

**Vivek Patidar**

B.Tech. IV Year

Department of Computer Science & Engineering

Faculty of Engineering

Medi-Caps University, Indore

**EXECUTIVE SUMMARY**

**Abstract**

Technological progress is accelerating day by day. Human-computer interaction is becoming more convenient as a result of speech-enabled virtual assistants that provide a wide range of services and can connect to smart environments, reducing physical interaction with the system. However, people face several issues while interacting with these assistants due to their numerous weaknesses, like a lack of understanding complex queries and not maintaining the personalized data of users, etc. As a consequence, the interaction is impaired, and in some cases, ineffective. For this reason, we have developed a virtual assistant using RPA (Robotic Process Automation), AI (Artificial Intelligence), Python, and ML (Machine Learning) that can eliminate these problems and do so with increased accuracy. Hence, this project proposes to create a virtual assistant that can make human-computer interaction as smooth and efficient as possible.

**Problem Statement**

In most cases, a user must manually manage many programs in order to execute a single activity. A user planning a trip, for example, should look up airport codes for neighboring airports and then search travel sites for tickets between airport combinations to get to their destination. There is a need for a system that can simply handle duties. We already have a number of virtual assistants, but we only utilize a few of them, and the ones we do use have certain drawbacks. One of these assistants' primary flaws is that they get overwhelmed with difficult questions and are unable to extract information from discussions. Hence, a virtual assistant who can interpret English in any accent and operate on a desktop system is required.

**Proposed Work**

Using sophisticated artificial intelligence (AI), robotic process automation (RPA), natural language processing, and machine learning, this virtual assistant pulls information and difficult data from interactions. Cognitive automation can aid if data is unstructured or traditional criteria aren't relevant. The phrase "cognitive automation" refers to a combination of RPA and data science approaches that excels at dealing with textual data. Virtual assistants should be able to model complicated task relationships and utilize these models to suggest user-friendly plans. With the high level of accuracy and precision required for VPA to be used on a daily basis, the program may be able to turn any speech with small changes or different accents into text.

**Conclusion**

We explored the design and execution of digital support in this project, "Virtual Personal Assistant Using RPA, AI, and Python." This project is more versatile and easier to add new features without compromising existing system operation since it is modular. It not only reacts to human commands but also gives responses based on the user's query or words, such as opening tasks and operations. In addition, the program should reduce any unnecessary manual work required in the user's daily chores. This is only the start; as AI and machine learning develop, our virtual assistant will get smarter and provide additional capabilities.

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **Chapters** | **Topic** | **Page No.** |
|  | Report Approval | ii |
|  | Declaration | iii |
|  | Certificate | iv |
|  | Acknowledgements | v |
|  | Executive Summary | vi |
|  | Table of Contents | viii |
|  | List of figures | x |
|  | List of tables | xii |
|  | Abbreviations | xiii |
| **Chapter 1** | Introduction | 1 |
|  | 1.1 Introduction | 1 |
|  | 1.2 Problem Statement | 5 |
| **Chapter 2** | Related Work | 9 |
| **Chapter 3** | System Analysis | 20 |
|  | 3.1 Requirements | 20 |
|  | 3.2 Modeling | 23 |
|  | 3.3 Functionality | 28 |
| **Chapter 4** | System Design | 29 |
|  | 4.1 System Structure | 29 |
|  | 4.2 Working Principles Used | 32 |
| **Chapter 5** | Implementation | 35 |
| **Chapter 6** | Results and Discussions | 45 |
| **Chapter 7** | Conclusion & Summary | 49 |
| **Chapter 8** | Future Scope | 50 |
|  | Appendix | 51 |
|  | A.1 Snippet of Code of Virtual Assistant | 51 |
|  | A.2 Statements and Declarations | 53 |
|  | Bibliography | 55 |

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **Figure No.** | **Figure Name** | **Page No.** |
| 1.1 | Virtual Assistant | 1 |
| 1.2 | Some tasks virtual assistants can perform | 2 |
| 1.3 | Smart Devices | 5 |
| 1.4 | Basic Overflow | 7 |
| 3.1 | Activity Diagram | 23 |
| 3.2 | Use Case Diagram | 24 |
| 3.3 | Sequence Diagram 1 | 25 |
| 3.4 | Sequence Diagram 2 | 26 |
| 3.5 | Component Diagram | 27 |
| 4.1 | Structure | 29 |
| 4.2 | System Architecture | 32 |
| 5.1 | Greeting 1 | 35 |
| 5.2 | Greeting 2 | 35 |
| 5.3 | Time | 36 |
| 5.4 | Google | 37 |
| 5.5 | YouTube | 37 |
| 5.6 | Music on Spotify | 38 |
| 5.7 | Amazon | 39 |
| 5.8 | Note taking on Google Keep | 39 |
| 5.9 | Instagram and Twitter | 40 |
| 5.10 | Weather | 41 |
| 5.11 | Rock, Paper and Scissors | 42 |
| 5.12 | Calculator | 43 |
| 5.13 | Exit | 44 |
| 6.1 | Google | 45 |
| 6.2 | Assistant Working | 46 |
| 6.3 | Result | 46 |
| 6.4 | Voice enabled VPA | 48 |

**LIST OF TABLES**

|  |  |  |
| --- | --- | --- |
| **Table No.** | **Table Name** | **Page No.** |
| 2.1 | Literature Review | 19 |

**ABBREVIATIONS**

|  |  |
| --- | --- |
| **Abbreviation** | **Full Form** |
| VPA | Virtual Personal Assistant |
| AI | Artificial Intelligence |
| IT | Information technology |
| HR | Human Resources |
| IDE | Integrated Development Environment |
| TV | Television |
| RPA | Robotic Process Automation |
| IVA | Intelligent Virtual Assistant |
| IPA | Intelligent Personal Assistant |
| NLP | Natural Language Processing |
| ASR | Automated Speech Recognition |
| HMM | Hidden Markov Model |
| MFCC | Mel Frequency Cepstral Coefficients |
| DTW | Dynamic Time Warping |
| DFT | Discrete Fourier Transform |
| DCT | Discrete Cosine Transform |
| LPC | Linear Predictive Coding |
| VUS | Virtual Users System |
| OS | Operating System |
| LAN | Local Area Network |
| API | Application Programming Interface |
| UI | User Interface |
| CE | Content Extraction |
| TTS | Text-to-Speech |
| ML | Machine Learning |

**CHAPTER 1**

**INTRODUCTION**

**1.1 INTRODUCTION**

People no longer rely on other humans for assistance or services. Because of the digitalization of the world, humans no longer need to rely on others for assistance; instead, they may rely on considerably more systematic and dependable equipment to meet their daily demands. VPA (Virtual Personal Assistant) has practically become a requirement. However, although it is reasonably efficient, it is not particularly useful and is not used by the user owing to its high mistake rate. Hence, the aim of our project is to create a VPA with a very low error rate.

As our world becomes more digital, virtual assistants powered by advanced AI are bridging the gap between the digital and human worlds. Consumers and corporations may get help with a variety of chores. But what is a virtual assistant, and what can it help you with?

**Figure 1.1:** Virtual Assistant.

A virtual assistant is a digital assistant that uses voice recognition features and language processing algorithms to recognize the voice commands of the user and perform relevant tasks as requested by the user.

Virtual assistants are extremely beneficial to the elderly, the blind and physically disabled, children, and others by ensuring that interacting with machines is no longer a barrier for them. Even blind people who can't see the gadget can use their voices to communicate with it.

Here are some of the basic tasks that can be done with the help of voice assistant:

* Reading Newspaper
* Interact with the user based on the questions asked
* Getting updates of mail
* Play games

  
**Figure 1.2:** Some tasks virtual assistants can perform.

* Play a music or video
* Setting a reminder and alarm
* Run any program or application
* Getting weather updates
* Make online purchases
* Control other smart devices (lights, locks, thermostats, vacuum cleaners, switches)

Virtual assistants may be beneficial in a variety of sectors, including IT helpdesk, home automation, HR-related chores, voice-based search, and so on. Voice-based search is the future of the next generation, with users relying heavily on voice assistants for all of their requirements.

A virtual assistant is capable of doing a variety of tasks. Users can use it to get basic answers and recommendations based on their profile, previous activity, and additional behavior. Moreover, our new version of the digital assistant can allow users to turn on the lights, make grocery lists, and turn off the heating when they are away from home.

Organizations can use this digital assistant in customer care to manage incoming contacts. It can also support a wide range of IT operations. Routine procedures like system upgrades, information management, and even transaction orders may be automated.

Moreover, it can be used in the automobile industry, making safety an increasingly important issue. Using voice commands instead of touch-tones is not only a convenience, it is being perceived by consumers as a safety necessity as the user doesn’t have to interact physically with the system and can focus on the driving while using just voice commands to interact.

This system is designed to be used efficiently on desktops and mobiles. Personal assistant software improves user productivity by managing routine tasks of the user and by providing information from online sources to the user. It is effortless to use and within seconds, it executes the command or query.

Voice searches have dominated over text search. Web searches conducted via mobile devices have only just overtaken those carried out using a computer and the analysts are already predicting that 75% of searches will be via voice by 2023. Virtual assistants are turning out to be smarter than ever. Allow your intelligent assistant to make email work for you. Detect intent, pick out important information, automate processes, deliver personalized responses and what not.

When someone works in an environment with which he/she is not familiar with, they often find it difficult to locate applications that they need, like a browser, any IDE or any other software. Most of the time, they will end up wasting hours of time, searching for the application alone. This results in unnecessary time wastage. Therefore, a voice enabled personal assistant will help automating this process. User is expected just to give a voice command, and the assistant will take care of the rest.

While indirect revenues for the carriers will be several folds. A few companies have started offering converging products in the VPA direction, e.g. Conita, WildFire, VoxSurf, VoiceGenie, and VoiceTel, and Mitel Networks, though one or two provide solutions for mobile carrier environments.

Also, the approach has some new methods that make this device unique, such as using as TV by using the data show or connecting the device with screen, watching TV and movies with translation language, chatting with anyone in any language, understanding body language and movements, and playing games with speech and gesture recognition; it also can be used to read facial and speech expressions.

This project was started on the premise that there is a sufficient amount of openly available data and information on the web that can be utilized to build a virtual assistant that has access to making intelligent decisions for routine user activities. In summary, we've developed an AI-powered virtual assistant that can execute all of these tasks and more without causing any disruptions.

**1.2 PROBLEM STATEMENT**

Usually, a user needs to manually manage multiple sets of applications to complete one task. For example, a user trying to make a travel plan needs to check for airport codes for nearby airports and then check travel sites for tickets between combinations of airports to reach the destination. There is a need for a system that can manage tasks effortlessly. We already have multiple virtual assistants but we hardly use them or there are some disadvantages associated with the ones we rarely use. For example, Apple’s Siri doesn’t maintain a knowledge database of its own and its understanding comes from the information captured in domain models and data models. Hence, the user doesn’t get a personalized experience. Similarly, Google’s assistant is known to consume a large amount of battery and it requires users to give access to a lot of data which raises privacy issues.



**Figure 1.3:** Smart Devices (a. Apple’s Homepod ; b. Amazon’s Echo ; c. Microsoft’s Cortana ; d. Google’s Home)

When a virtual assistant is not able to answer questions accurately, it’s because it lacks the proper context or doesn’t understand the intent of the question. Its ability to answer questions relevantly only happens with rigorous optimization, involving both humans and machine learning. Continuously ensuring solid quality control strategies will also help manage the risk of the virtual assistant learning undesired bad behaviors. They require a large amount of information to be fed in order for it to work efficiently.

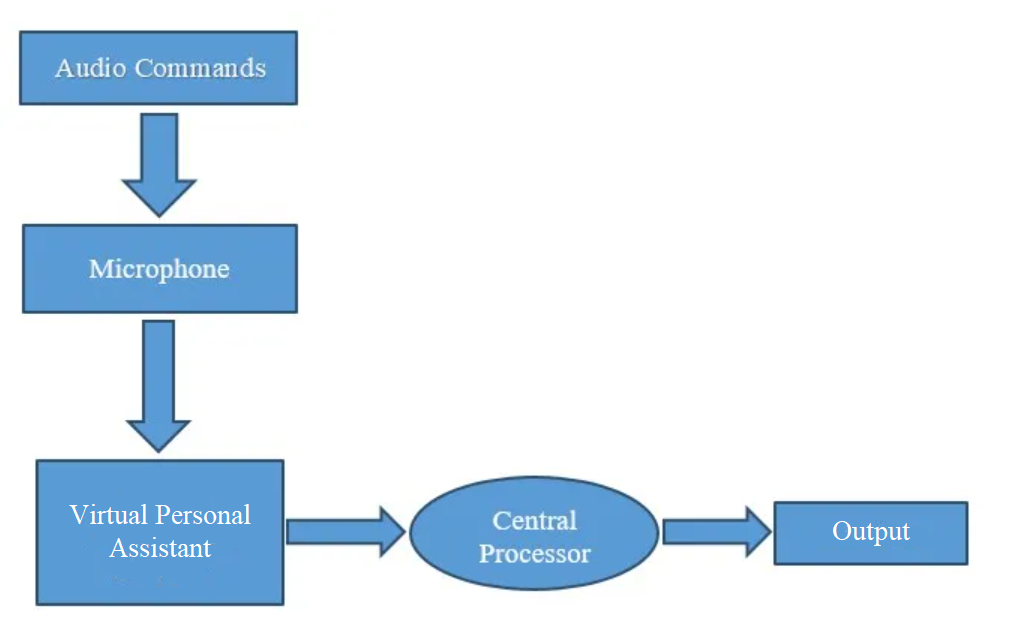
Virtual assistants should be able to model complex task dependencies and use these models to recommend optimized plans for the user. It needs to be tested for finding optimum paths when a task has multiple sub-tasks and each sub-task can have its own sub-tasks. In such a case there can be multiple solutions to paths, and it should be able to consider user preferences, other active tasks, priorities in order to recommend a particular plan.

Moreover, people have so many issues with these assistants. These systems can understand English phrases but they fail to recognize our accent. Also, they are easier to use on mobile devices than desktop systems. And one of the major drawbacks of these assistants is that they get overwhelmed during a complex query and can’t extract information from conversations. There is a need for a virtual assistant that can understand English in any accent and work on a desktop system. When a virtual assistant is not able to answer questions accurately, it’s because it lacks the proper context or doesn’t understand the intent of the question. Its ability to answer questions relevantly only happens with rigorous optimization, involving both humans and machine learning.

**1.3 PROPOSED WORK**

Virtual assistants should be able to model complex task dependencies and use these models to recommend optimized plans for the user. It needs to be tested for finding optimum paths when a task has multiple sub-tasks and each sub-task can have its own sub-tasks. In such a case, there can be multiple solutions to paths, and it should be able to consider user preferences, other active tasks, and priorities in order to recommend a particular plan.

Main objective of building a personal assistant is using semantic data sources available on the web, user generated content and providing knowledge from knowledge databases. One of the main purposes of an intelligent virtual assistant is to answer questions that users may have. This may be done in a business environment, for example, on the business website, with a chat interface. On the mobile platform, the virtual personal assistant is available as a call-button operated service where a voice asks the user for a command or query and responds accordingly.

****

**Figure 1.4:** Basic Workflow

One of the main advantages of voice search is its rapidity. In fact, voice is reputed to be almost four times faster than a written search. We can write about 40 words per minute, whereas we are capable of speaking around 150 during the same period of time. In this respect, the ability of personal assistants to accurately recognize spoken words is a prerequisite for them to be adopted by consumers.

Virtual assistants are the most advanced kind of robotic automation. Our virtual assistant gathers information from a variety of sources and contextualizes it, learning from each encounter. The virtual assistant can process everything that is said or written and utilize it to generate an accurate answer using advanced language processing. Using AI and machine learning, our virtual assistant can handle various tasks and complicated queries. It uses neural networking to acquire insight into one's preferences based on prior selections and data. In this way, when the user interacts with the virtual assistant, it becomes a unique experience that meets their needs.

This virtual assistant extracts information and complicated data from interactions using advanced artificial intelligence (AI), robotic process automation (RPA), natural language processing, and machine learning. RPA is a smart software application that excels in automating repetitive and rule-based processes using highly organized data. However, RPA can also get confused during the time of a complex query. If data is more unstructured or conventional criteria aren't applicable, cognitive automation can help. The term "cognitive automation" refers to a mix of RPA and data science methodologies that is particularly effective when dealing with textual information. With a speech and text-based user interface, our virtual assistant takes this to the next level. It is capable of extracting information and complicated facts from conversations and comprehending them appropriately.

The main goal of this project is to improve the accuracy of the voice-to-text software. That is, the software will potentially be able to translate any speech with minor modulations or varied accents into text with the high degree of accuracy and precision required for VPA’s day-to-day use.

**CHAPTER 2**

**RELATED WORK**

A. Sudhakar Reddy M., Vyshnavi, C. Raju Kumar, and Saumya, "VIRTUAL ASSISTANT USING ARTIFICIAL INTELLIGENCE ", International Journal of Emerging Technologies and Innovative Research (www.jetir.org), ISSN:2349-5162, Vol.7, Issue 3, page no. 1116-1119, March-2020 [1]. In this paper the authors discussed whether an intelligent virtual assistant (IVA) or intelligent personal assistant (IPA) may be a software agent which will perform tasks or services for privately supported commands or questions. Authors quoted that “sometimes the term "chatbot" is employed to ask for virtual assistants generally or specifically accessed by online chat or in some cases, online chat programs are exclusively for entertainment purposes.” Some virtual assistants are ready to interpret human speech and respond via synthesized voices. Users can ask their assistants questions, control home automation devices and media playback via voice, and manage other basic tasks like email, todo lists, and calendars with verbal commands.

The paper described a new emerging service for mobile users. The Virtual Personal Assistance provides an intelligent computer secretarial service for mobile professionals. The new service is based on convergence of the internet, speech recognition technology and mobile technologies. The VPA minimizes the interruption of the user, improves the utilization of his time, and provides a single point of communication for all his messages, contacts, schedule, and source of information. The paper proposes as well a decision structure for call screening, as well as handling requests for meetings and appointments. The system initially targets lawyers, doctors, sales personnel, small offices, maintenance crews, etc. However, it is expected to become a standard feature for millions of other users.

Though the authors mentioned improving user efficiency and time consumption, they were unable to provide any proof for their claim. According to them, this VPA overcomes many drawbacks, but there aren't any particular disadvantages listed in the paper. In conclusion, the authors talked about the errors and accuracy/precision problems of currently existing assistants but failed to give a proper solution. However, the research done on how NLP and AI can help these assistants is very useful [1].

George Terzopoulos, Maya Satratzemi, “Voice Assistants and Smart Speakers in Everyday Life and in Education”, Informatics in Education, v19 n3 p473-490 2020 [2]. In this paper the author discussed how in recent years, Artificial Intelligence (AI) has shown significant progress and its potential is growing. An application area of AI is Natural Language Processing (NLP). Voice assistants incorporate AI by using cloud computing and can communicate with the users in natural language. Voice assistants are easy to use and thus there are millions of devices that incorporate them in households nowadays. Most common devices with voice assistants are smart speakers and they have just started to be used in schools and universities. The purpose of this paper is to study how voice assistants and smart speakers are used in everyday life and whether there is potential in order for them to be used for educational purposes.

The paper talked about immersive learning technologies that have the ability to update the existing education system. Virtual reality, augmented reality and voice assistants can provide new learning experiences. In this paper, research regarding the integration of AI voice assistants in education is presented. Research on this topic is limited since voice assistants and smart speakers are now gaining popularity. Findings presented in this paper will hopefully inspire other researchers to further investigate this topic. Smart speakers and voice assistants will be at the center of interest in coming years as they enter the everyday life of households. The ways they can be used efficiently in the learning process is the subject of research as there are many challenges. One of these challenges is the lack of many languages as voice assistants do not speak all languages. In addition, voice assistants do not have many of the appropriate security measures and protection filters that can be used in class by students. Teachers need to be trained and motivated about the usefulness of these devices in order to adopt them in their class. Although in most cases positive results have been reported regarding students and teachers, results are limited, incomplete and unorganized. As a conclusion, the role of these devices and their use in the classroom are still at an early stage of research and more studies need to address this topic.

In conclusion, this paper explains the future scope of virtual personal assistants in external devices and everyday life. This paper shows us how we can use IoT to connect these assistants with various devices, like Amazon has done with Alexa, and create a smart speaker. Hence, this led to the idea of connecting our devices with automobiles [2].

Regina Gubareva, Rui Pedro Lopes, “Virtual Assistants for Learning: A Systematic Literature Review”, CSEDU20-RP-93, May-2020 [3]. In this paper the author discussed a problem of students’ motivation, engagement and declining interest in the learning process that has always existed, contributing to increasing failures and dropouts. This is particularly important among first year students. Freshmen often have difficulties with time management, how to prioritize tasks, and how to study at the university. Because of the increasing number of higher education students, it is impossible to provide individual tutoring and support to every student, to help them manage this first-year indefiniteness and later difficulties. Recent developments in the area of information technology, software engineering, artificial intelligence, machine learning, and big data creates the opportunity for personalized, flexible and adaptable learning environments, accessible anytime, anywhere. One such example is a virtual assistant, a tool that provides assistance to usually boring or repetitive daily activities. In education, a virtual assistant can help organize the study process, manage time, increase motivation and engagement in the study process. This paper performs a systematic literature review of the use of virtual assistants in higher education. It focuses on the technology that powers them, their features and their impact in the learning process, motivation and productivity, according to the authors.

In this paper the author explored the research in the field of virtual assistants in higher education. The review provides an array of uses and techniques for developing virtual assistants and how they can be used for education support. Student support can be categorized depending on the role of virtual assistant in four types: digital tutor, digital secretary, the motivator agent, the mentor agent. The tutor’s role is to help students through the learning process The digital secretary helps organize the learning process, solving administrative issues, and reminding deadlines. The motivator helps integrating in social environments, to help overcome stress and anxiety, remind the main aim and motivate students. Finally, the mentor gives a general description of how to achieve the solution and description of exactly which problem-solving action steps should be taken. Virtual assistants are becoming popular and useful technology, with a variety of advantages, contributing to automation of tasks and providing support for students in time-management, access to information and communication facilitation. The technology is still in its infancy. There are many aspects that are necessary to improve to make virtual assistants effective in student motivation and engagement.

This paper showed how virtual assistants are versatile in performing different tasks and how they can be used in different fields. The author described how these assistants can be used to motivate students by taking care of their tasks and, hence, motivating us to include as many tasks as we can through our assistant. This also shows how VPA’s have unlimited scope and they will keep improving and assisting in different areas forever [3].

Emad S. Othman. “Voice Controlled Personal Assistant Using Raspberry Pi”. International Journal of Scientific and Engineering Research Volume 8, Issue 11, November-2017 [4]. The purpose of this paper was to illustrate the implementation of a Voice Command System as an Intelligent Personal Assistant (IPA) that can perform numerous tasks or services for an individual. These tasks or services are based on user input, location awareness, and the ability to access information from a variety of online sources (such as weather or traffic conditions, news, stock prices, user schedules, retail prices, telling time, local traffic, travel assistant, events, notification from social applications plus one can ask questions to the system, invoke its machine learning otherwise get it from Wikipedia… etc…..). Using Raspberry Pi as a main hardware to implement this model which works on the primary input of a user’s voice. Using voice as an input to convert into text using a speech to text engine. The text hence produced was used for query processing and fetching relevant information. When the information was fetched, it was then converted to speech using text to speech conversion and the relevant output to the user was given. Additionally, some extra modules were also implemented which worked on the concept of keyword matching. It can help the visually impaired to connect with the world by giving them access to Wikipedia, Calculator, Email and Music all through their voice. This model can also keep people secure as it can be used as a surveillance system which captures the voice of the person standing at the door and similarity checking. Also it can be a source of entertainment and information for blind/visually impaired. This model will interact with other systems by means of IOT, thus providing a fully automated system. Many experiments and results were accomplished and documented.

In this paper, the author introduced the idea and rationale behind the Voice Command System, the flaws in the current system and the way of resolving those flaws and laid out the system architecture of the presented Voice Command System. Many modules are of open-source systems and have customized those modules according to the presented system. This helps get the best performance from the system in terms of space time complexity. The Voice Command System has an enormous scope in the future. Like Siri, Google Now and Cortana have become popular in the mobile industry. This makes the transition smooth to a complete voice command system. Additionally, this also paves way for a Connected Home using Internet of Things, voice command system and computer vision.

This project looks promising and the research done and the modules that the author worked on gave us insight on the different activities that we can let our assistant perform. However, the Raspberry Pi has numerous disadvantages, such as it is not able to work on desktop systems, its graphics processor is missing, etc. [4].

V. Radha and C. Vimala, “A review on speech recognition challenges and approaches,” doaj. org, vol. 2, no. 1, pp. 1–7, 2012 [5]. Speech technology and systems in human computer interaction have witnessed a stable and remarkable advancement over the last two decades. Today, speech technologies are commercially available for an unlimited but interesting range of tasks. These technologies enable machines to respond correctly and reliably to human voices, and provide useful and valuable services. Recent research concentrates on developing systems that would be much more robust against variability in environment, speaker and language. Hence today’s researches mainly focus on ASR systems with a large vocabulary that support speaker independent operation with continuous speech in different languages. This paper gives an overview of the speech recognition system and its recent progress. The primary objective of this paper is to compare and summarize some of the well-known methods used in various stages of speech recognition systems.

Speech recognition has been in development for more than 50 years, and has been entertained as an alternative access method for individuals with disabilities for almost as long. In this paper, the fundamentals of speech recognition are discussed and its recent progress is investigated. The various approaches available for developing an ASR system are clearly explained with its merits and demerits. The performance of the ASR system based on the adopted feature extraction technique and the speech recognition approach for the particular language is compared in this paper. In recent years, the need for speech recognition research based on large vocabulary speaker independent continuous speech has highly increased. Based on the review, the potent advantage HMM approach along with MFCC features is more suitable for these requirements and offers good recognition results. These techniques will enable us to create increasingly powerful systems, deployable on a worldwide basis in future.

Automated speech recognition (ASR) is a technology that allows users of information systems to speak entries rather than punching numbers on a keypad. ASR is used primarily to provide information and to forward telephone calls.

The authors explained that the most general mode of communication among human beings is speech and hence having a speech-based assistant will solve most of our problems. As this is the utmost technique, human beings would be identical to utilize speech to interrelate with machines too. Because of this, autonomous speech identification has got a lot of reputation. Most techniques for speech recognition are like DTW (Dynamic Time Warping), HMM (Hidden Markov Model) [5].

Bassam A. & Raja N. (2010) “Arabic speech recognition using Hidden Markov Model Toolkit (HTK)”. 2. 557 - 562. 10.1109/ITSIM.2010.5561391 [6]. In this paper the author discussed the development and implementation of an Arabic automatic speech recognition engine. The engine can recognize both continuous speech and isolated words. The system was developed using the Hidden Markov Model Toolkit. First, an Arabic dictionary was built by composing the words to its phones. Next, Mel Frequency Cepstral Coefficients (MFCC) of the speech samples are derived to extract the speech feature vectors. Then, the training of the engine based on triphones is developed to estimate the parameters for a Hidden Markov Model. To test the engine, the database consisting of speech utterance from thirteen Arabian native speakers is used which is divided into ten speaker-dependent and three speaker-independent samples.

This paper talked about statements and speech which are most significant. Communication between humans and machines was done through analog signals which were converted by speech signals to digital waves. This technology is massively utilized, it has limitless uses and permits machines to reply appropriately and consistently to the user's voice, and also offers useful and appreciated facilities.

Hidden Markov Models (HMMs) are a class of probabilistic graphical models that allow us to predict a sequence of unknown (hidden) variables from a set of observed variables. A simple example of an HMM is predicting the weather (hidden variable) based on the type of clothes that someone wears (observed). The MFCC feature extraction technique basically includes windowing the signal, applying the DFT, taking the log of the magnitude, and then warping the frequencies on a Mel scale, followed by applying the inverse DCT. The detailed description of various steps involved in the MFCC feature extraction is explained below.

This paper talked about sentence correction, word correction and word accuracy using speech recognition and hidden markov models. The engine was checked for continuous sentences as well as isolated words. The system was tested by thirteen different Arabic native speakers and the experimental results showed that the overall system performance was 90.62%, 98.01% and 97.99% for sentence correction, word correction and word accuracy respectively. This paper helped us improve the accuracy of our assistant [6].

Elshafei, M., Virtual personal assistant (VPA) for mobile users. Mitel Networks (2000–2002) [7]. In this paper the author discussed how Virtual Personal Assistant (VPA) is the next generation of carrier services for mobile users. The author talked about how VPA is believed to be the intelligent evolution of services to meet the ever-increasing demand by the mobile professionals for mobility and connectivity and this new generation of services will allow mobile users to remotely access and manage information using speech recognition technology over telephones. VPA responds to conversational voice commands and delivers a single point of contact that seamlessly engages a wide range of information. The VPA controls the telephone calls, manages the personal activities through calendar, enables the user to access his task manager via voice interface, and includes all the functions of Unified Messaging. The VPA enables the user to optimize the user resources (time, cost), enhance his/her overall productivity, and minimize the interruptions to his regular workflow. The Virtual Personal Assistant (VPA) will enable the user to efficiently handle increasing demand for telephone calls, messages, meetings and other activities. The paper provides an overview of the VPA applications, and the expected features and future trends. The paper proposes as well a unified decision model based on a quantitative assessment of the importance of the requests and the availability of the user.

The paper described a new emerging service for mobile users. The Virtual Personal Assistance provides an intelligent computer secretarial service for mobile professionals. The new service is based on convergence of the internet, speech recognition technology and mobile technologies. The VPA minimizes the interruption of the user, improves the utilization of his time, and provides a single point of communication for all his messages, contacts, schedule, and source of information. The paper proposes as well a decision structure for call screening, as well as handling requests for meetings and appointments. The system initially targets lawyers, doctors, sales personnel, small offices, maintenance crews, etc. However, it is expected to become a standard feature for millions of other users.

This paper gives a generic overview of what VPAs were imagined to be 20 years ago. This is a very complex process, and it does tell us about how VPA was a dream back then. The author has listed out several basic services a VPA could perform that even any assistant or chatbot could do for you. But this could be thought of as the very start of how VPA’s were structured and how the idea of a VPA was generated. Even though it’s a very old and basic paper, it still tells us how virtual assistants have improved over time and how every piece of research has been helpful towards the intelligent and smart assistants we have now [7].

B. S. Atal and L. R. Rabiner, “A pattern recognition approach to voiced unvoiced-silence classification with applications to speech recognition,” Acoustics, Speech and Signal Processing, IEEE Transactions on, vol. 24, no. 3, pp. 201–212, 1976 [8]. In this paper the author talked about speech analysis, the voiced-unvoiced decision is usually performed in conjunction with pitch analysis. The linking of voiced-unvoiced (V-UV) decision to pitch analysis not only results in unnecessary complexity, but makes it difficult to classify short speech segments which are less than a few pitch periods in duration. In this paper, we describe a pattern recognition approach for deciding whether a given segment of a speech signal should be classified as voiced speech, unvoiced speech, or silence, based on measurements made on the signal. In this method, five different measurements are made on the speech segment to be classified. The measured parameters are the zero-crossing rate, the speech energy, the correlation between adjacent speech samples, the first predictor coefficient from a 12-pole linear predictive coding (LPC) analysis, and the energy in the prediction error. The speech segment is assigned to a particular class based on a minimum-distance rule obtained under the assumption that the measured parameters are distributed according to the multidimensional Gaussian probability density function. The means and covariances for the Gaussian distribution are determined from manually classified speech data included in a training set. The method has been found to provide reliable classification with speech segments as short as 10 ms and has been used for both speech analysis-synthesis and recognition applications. A simple nonlinear smoothing algorithm is described to provide a smooth 3-level contour of an utterance for use in speech recognition applications. Quantitative results and several examples illustrating the performance of the method are included in the paper.

A fairly general framework based on a pattern recognition approach to VUS classification has been described in which a set of measurements are made on the interval being classified, and a minimum non-Euclidean distance measure is used to Select the appropriate class. Almost any set of measurements can be used so long as there is some physical basis for assuming that the measurements are capable of reliably distinguishing between these three classes.

The authors explained regarding speech analysis, and the result is regularly completed in combination with pitch analysis. The research described a pattern recognition technique for determining whether a given slice of a speech signal should be categorized as voiced speech, unvoiced speech, or silence, depending on dimensions finished on signal. The main restriction of the technique is the requirement to exercise the algorithm on the exact set of dimensions picked, and for the specific recording circumstances [8].

**GOOGLE ASSISTANT**

Google Assistant is a virtual assistant software application developed by Google that is primarily available on mobile and home automation devices. Based on artificial intelligence, Google Assistant can engage in two-way conversations, unlike the company's previous virtual assistant, Google Now. Google Assistant was released on May 18, 2016.

**PROS:**

Well, as modern technology is upgrading day by day, it is making our days easier. So, the Google Assistant is also one of the factors that makes our life easier. It can control your devices and your smart home. So, technically speaking you can access your calendars and make your schedule. Voice command can be given to search the keyword you are searching for and read your content. You can even call your contact numbers and answer your call with a voice command. Google Assistant can sync your preferred searching terms across devices such as Computers, Smartphones, Google Home, and other platforms. So, it helps to find your query quickly and learn them by AI and makes it easier to perform your task quicker across the platforms. You can even sync your search history as well. Google Assistant can learn your preferences and it will help to save your time when switching across the devices.

**CONS:**

Although Google Assistant has man-like voiceovers, it still is not clear yet. We can hear some cracks and unpleasant sound breaks while performing the action. We feel like we’re interacting with ROBOTS, not the assistant. Since the objective of Goole was to mimic interaction with the real person, it still lacks that experience and does not feel like interacting with the real person. Google invented the Duplex, an extension that assists Google assistants to carry out natural conversation by mimicking human voice. The manner is similar to robocalling. The assistant can complete tasks like calling your loved ones, turn off fans, and many more. It can even interact like humans, such as giving replies like “hmm” and “uh”, “mhm” and “gotcha”. But there have been many unethical and privacy concerns over critics. Since we are giving our information to the assistant it might later turn out to be a sci-fi horror movie. Privacy advocates even raised concerns about the millions of vocal samples being recorded and used as feedback for virtual assistants. It makes AI even smarter and disadvantageous for us.

**APPLE’S SIRI**

Siri is a virtual assistant that is part of Apple Inc.'s iOS, iPadOS, watchOS, macOS, tvOS, and audioOS operating systems. Siri is a personal assistant software that interfaces with the user through voice interface, recognizes commands and acts on them. It learns to adapt to the user's speech and thus improves voice recognition over time. It also tries to converse with the user when it does not identify the user request. It integrates with calendar, contacts and music library applications on the device and also integrates with GPS and camera on the device. It uses location, temporal, social and task-based contexts, to personalize the agent behavior specifically to the user at a given point of time. Siri was released on October 4, 2011.

**PROS:**

Siri is a master of record-keeping, it helps people from maintaining copious amounts of documentation, and streamlines the process of note-taking and documenting. Siri is mostly used in emailing and calling among the general population. The problem-solving abilities of Siri are also top notch, even when it couldn’t give the exact answers yet it still helps. Out of the box, Siri’s language is English. But, if you don’t speak English fluently or simply want to practice a language you are currently learning, Siri can be a great help, as Siri comes with non-English language options. Siri is not available only for iPhone, iPad and iPod touch. You can also use Siri on Apple Watch, Apple TV (generation 4) and your Mac. She can do many things on all these devices as well. Take Mac as an example. You can ask Siri how much free space you have on your [MacBook](https://tech.trade/refurbished-macbooks), search for files on your Mac, search the web and more.

**CONS:**

Siri has a lack of function. Either she’s having trouble connecting, or she’s turning your dictation into gibberish, or she’s bringing the web results you didn’t ask for. Apple has even been sued for overstating its claims of Siri’s abilities. Also, to date, Apple’s website continues to refer to Siri as “beta only”. As in, beta testing, a stage usually reserved for unfinished, unshipped products. It’s like they’re leaving the designation in place as a way of writing off complaints. Their answer to the functionality lawsuit? If you hated Siri so much, you should have [brought the phone back](https://bgr.com/2012/05/17/apple-siri-lawsuit-beta-performance/). Siri’s data collecting is vague and not advanced enough. The biggest disadvantage of Siri is that it does not maintain a knowledge database of its own and its understanding comes from the information captured in domain models and data models.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **S. No.** | **Name** | **Author** | **Date** | **Source** |
| 1. | Virtual Assistance using Artificial Intelligence. | A. Sudhakar Reddy M., Vyshnavi, C. Raju Kumar, and Saumya. | March, 2020 | JETIR |
| 2. | Voice Assistants and Smart Speakers in Everyday Life and in Education. | George Terzopoulos, Maya Satratzemi. | 2020 | Informatics in Education |
| 3. | Virtual Assistants for Learning: A Systematic Literature Review. | Regina Gubareva, Rui Pedro Lopes. | May, 2020 | SciTePress |
| 4. | Voice Controlled Personal Assistant Using Raspberry Pi. | Emad S. Othman. | November, 2017 | IJSER |
| 5. | A review on speech recognition challenges and approaches. | V. Radha and C. Vimala. | 2012 | DOAJ |
| 6. | Arabic speech recognition using Hidden Markov Model Toolkit. | Bassam A., Raja N. | 2010 | IEEE |
| 7. | Virtual personal assistant for mobile users. | Elshafei, M. | 2000-2002 | IJIRT |
| 8. | A pattern recognition approach to voiced unvoiced-silence classification with applications to speech recognition. | B. S. Atal and L. R. Rabiner. | 1976 | IEEE |
| 9. | Virtual Assistant using Python. | Jatu Naazneen, Abdul Gaffar, Abhijit Palse. | 2018-2019 | DTSS College |
| 10. | Google Assistant | Google | May 18, 2016 | Google |
| 11. | Siri | Apple | October 4, 2011 | Apple |

**Table 2.1:** Literature Review

**CHAPTER 3**

**SYSTEM ANALYSIS**

**3.1 REQUIREMENTS**

Personal assistant software is required to act as an interface into the digital world by understanding user requests or commands and then translating into actions or recommendations based on the agent's understanding of the world. VPA focuses on relieving the user of entering text input and using voice as primary means of user input. It then applies voice recognition algorithms to this input and records the input. It then uses this input to call one of the personal information management applications such as task list or calendar to record a new entry or to search about it on search engines like Google. Focus is on capturing the user input through voice, recognizing the input and then executing the tasks if the agent understands the task. Software takes this input in natural language, and so makes it easier for the user to input what he or she desires to be done. Voice recognition software enables hands free use of the applications, letting users to query or command the agent through voice interface. This helps users to have access to the assistant while performing other tasks and thus enhances the value of the system itself. VPA also has ubiquitous connectivity through Wi-Fi or LAN connection, enabling distributed applications that can leverage other APIs exposed on the web without a need to store them locally.

**3.1.1 FEASIBILITY STUDY**

1. Technical Feasibility : It includes finding out technologies for the project, both hardware and software. For a virtual assistant, the user must have a microphone to convey their messages and a speaker to listen when the system speaks. These are very cheap nowadays and everyone generally possesses them. Besides, the system needs an internet connection for some queries that are web related so, make sure you have a steady internet connection. It is also not an issue in this era where almost every home or office has Wi-Fi.
2. Operational Feasibility : It is the ease and simplicity of operation of the proposed system. System does not require any special skill set for users to operate it. In fact, it is designed to be used by almost everyone. Kids who still don’t know how to write can read out problems for the system and get answers.
3. Economical Feasibility : Here, we find the total cost and benefit of the proposed system over the current system. For this project, the main cost is documentation cost. Users also would have to pay for a microphone and speakers but generally every desktop and mobile system has it in-built so it shouldn’t be any problem. As far as maintenance is concerned, the assistant won’t cost too much.
4. Organizational Feasibility : This shows the management and organizational structure of the project. This project is built by a team of two persons. The management tasks are all to be carried out by them only. That won’t create any management issues and will increase the feasibility of the project
5. Cultural Feasibility : It deals with compatibility of the project with the cultural environment. Virtual assistant is built in accordance with the general culture without undermining any beliefs.

This project is technically feasible with no external hardware requirements. Also, it is simple in operation and does not cost training or repairs. Overall feasibility study of the project reveals that the goals of the proposed system are achievable.

**3.1.2 HARDWARE AND SOFTWARE REQUIREMENTS**

The software is designed to be light-weighted so that it doesn’t be a burden on the machine running it. This system is being built keeping in mind the generally available hardware and software compatibility. Here are the minimum hardware and software requirements for virtual assistants.

1. HARDWARE

* Pentium-pro processor or later.
* RAM 512 MB or more.

1. SOFTWARE

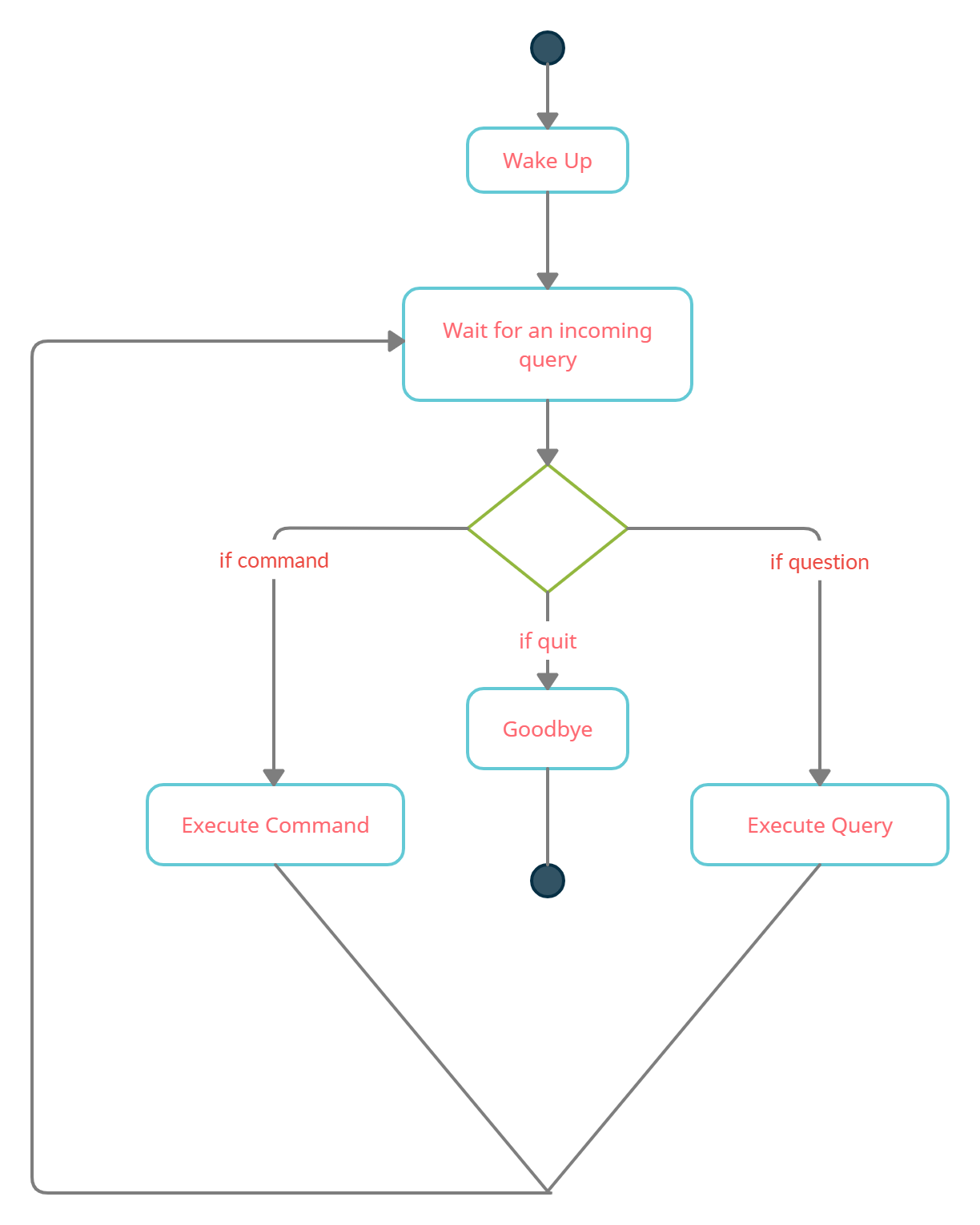
* Windows 7 (32-bit) or above.
* Python 2.7 or later.
* Chrome Driver.
* Selenium Web Automation.
* SQLite.

1. NON-FUNCTIONAL REQUIREMENTS

* The system ensures safety, security, and usability, which are observable during operation (at run time).
* The system is adaptable to different situations.
* The system has a good and compact UI with responsive interface.
* The project is light on resources.

**3.2 MODELING**

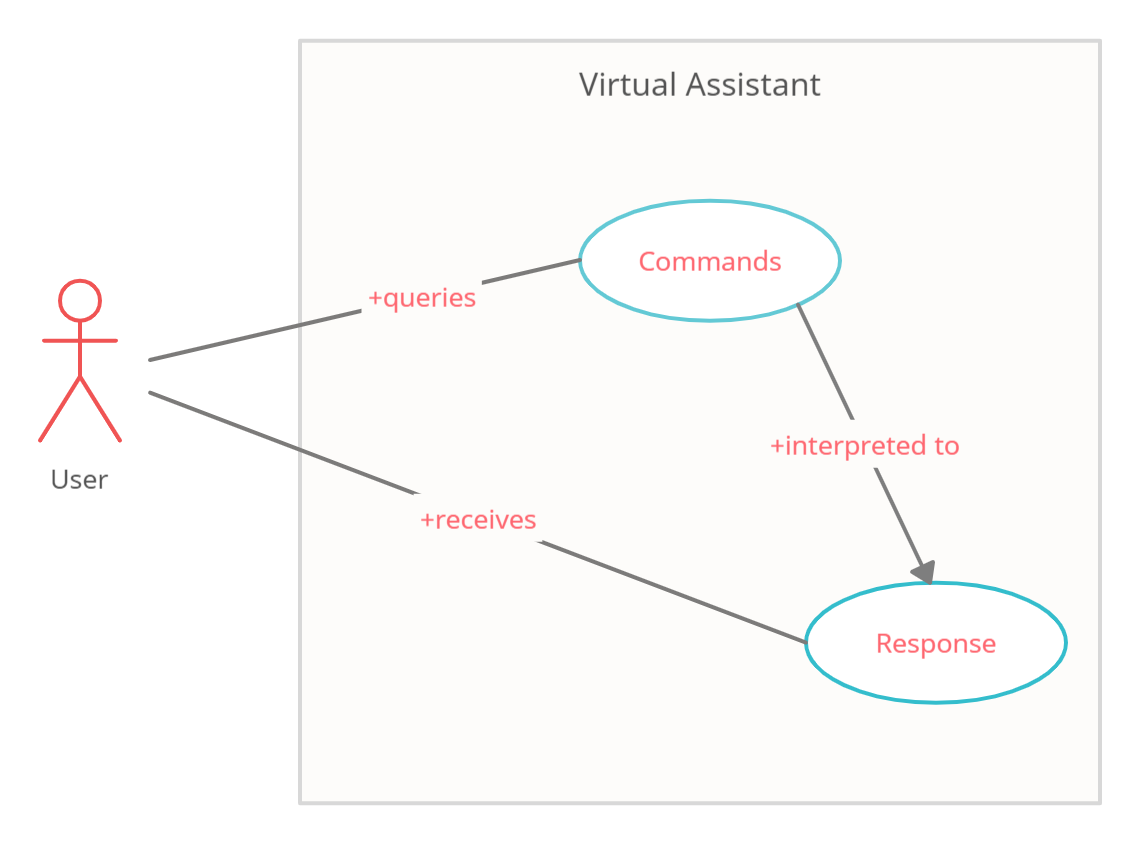
**3.2.1 ACTIVITY DIAGRAM**



**Figure 3.1:** Activity Diagram.

Initially, the system is in idle mode. As it receives any wake up call it begins execution. The received command is identified whether it is a questionnaire or a task to be performed. Specific action is taken accordingly. After the Question is being answered or the task is being performed, the system waits for another command. This loop continues unless it receives a quit command. At that moment, it goes back to sleep.

**3.2.2 USE CASE DIAGRAM**

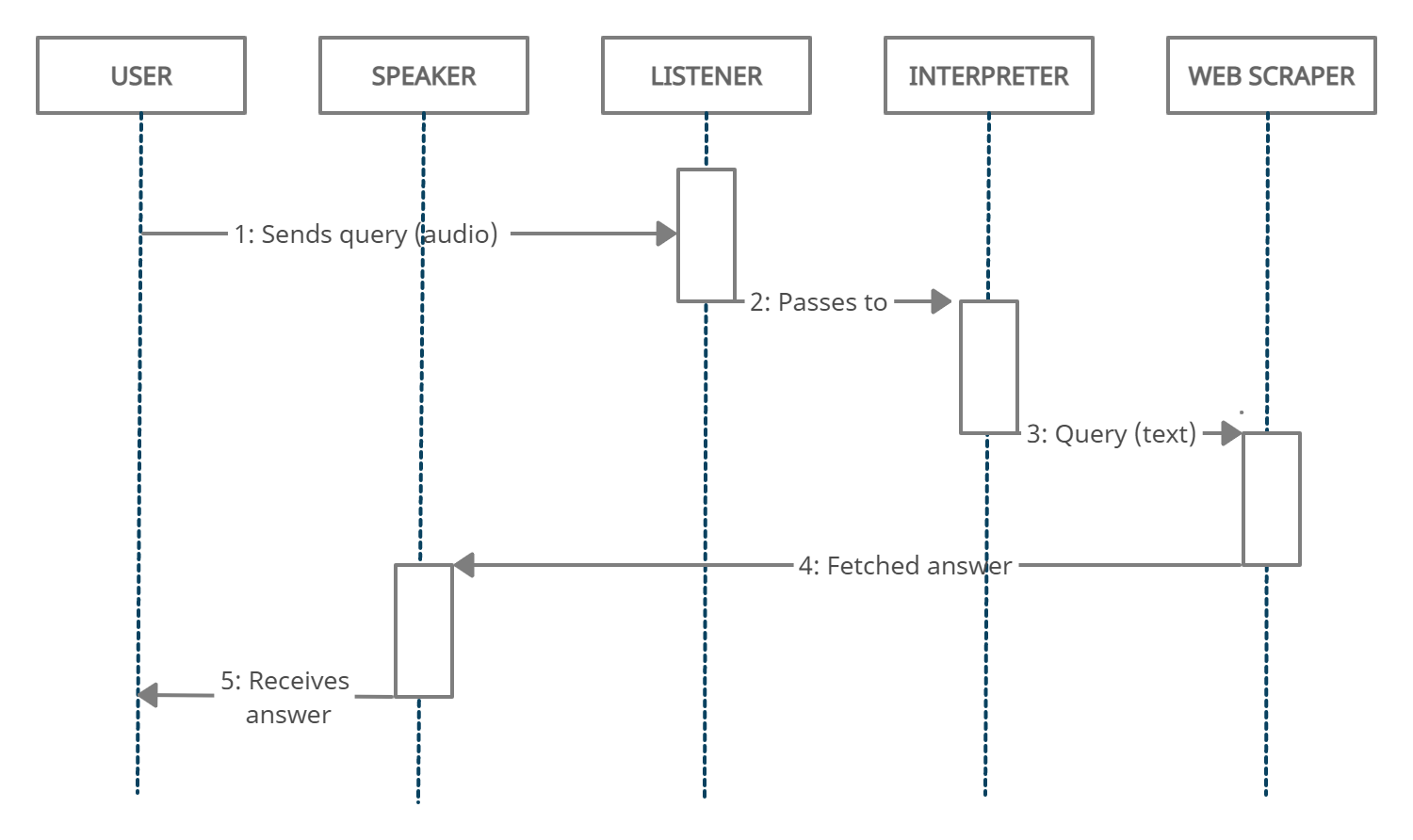


**Figure 3.2:** Use Case Diagram.

In this project there is only one user. The user queries command to the system. System then interprets it and fetches an answer. The response is sent back to the user.

**3.2.3 SEQUENCE DIAGRAM**

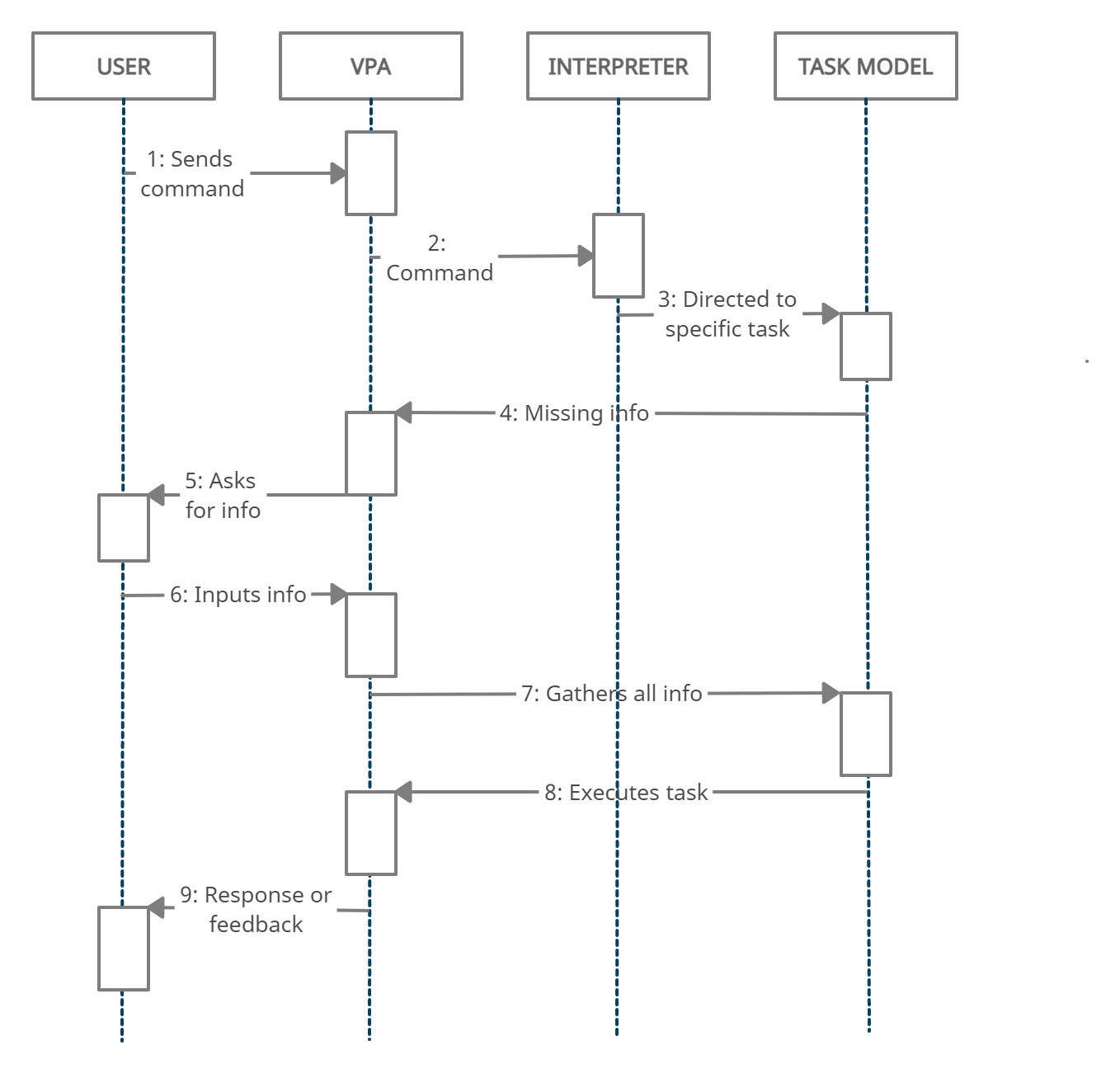
1. Sequence Diagram for Query-Response -

****

**Figure 3.3:** Sequence Diagram 1.

The above sequence diagram shows how an answer asked by the user is being fetched from the internet. The audio query is interpreted and sent to Web scraper. The web scraper searches and finds the answer. It is then sent back to the speaker, where it speaks the answer to the user.

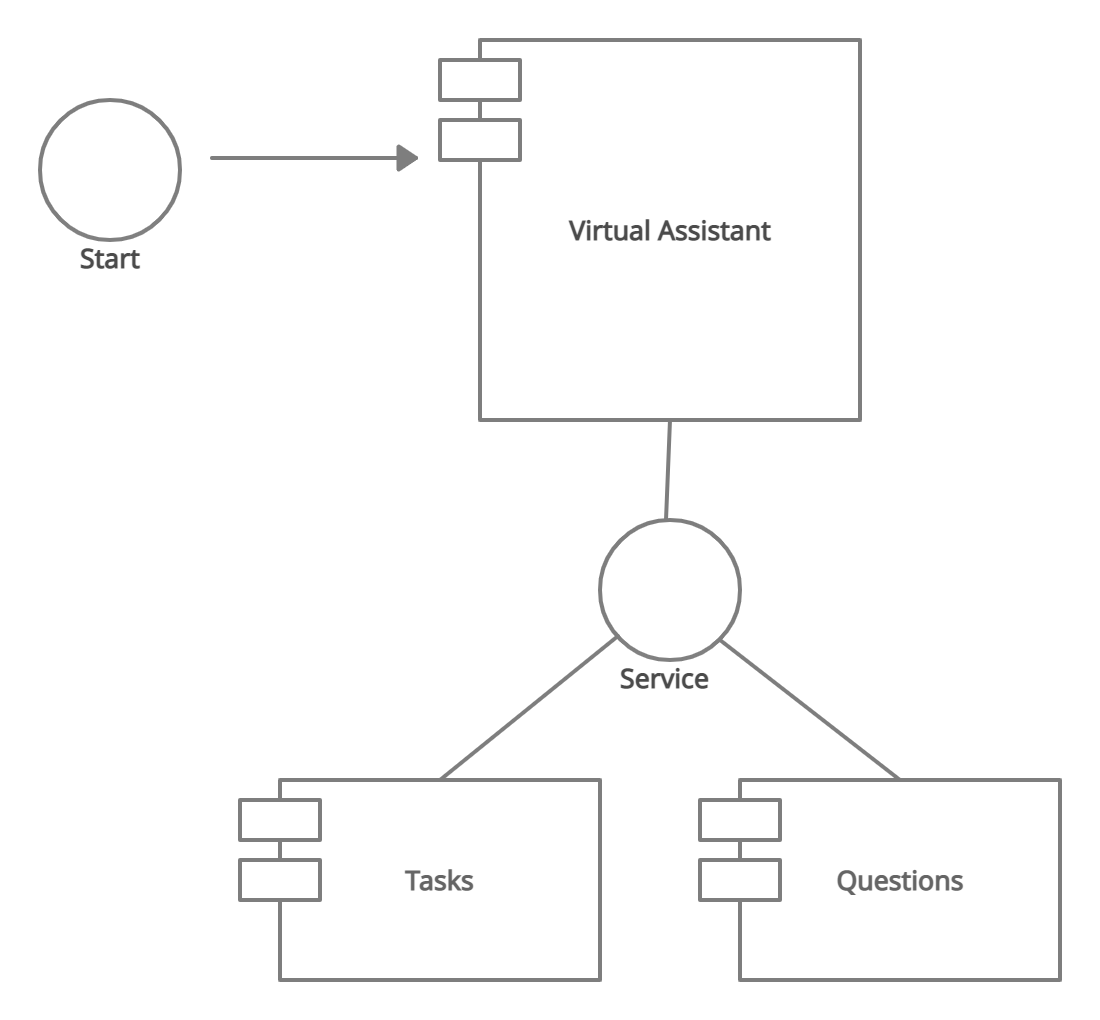
1. Sequence Diagram for task execution -

****

**Figure 3.4:** Sequence Diagram 2.

The user sends commands to the virtual assistant in audio form. The command is passed to the interpreter. It identifies what the user has asked and directs it to the task executor. If the task is missing some info, the virtual assistant asks the user back about it. The received information is sent back to task and it is accomplished. After execution feedback is sent back to the user.

**3.2.4 COMPONENT DIAGRAM**

****

**Figure 3.5:** Component Diagram.

The main component here is the Virtual Assistant. It provides two main services, executing tasks or answering your questions.

**3.3 FUNCTIONALITY**

The proposed system will have the following functionality:

(a) The assistant will be started by the user either manually or by voice activation (the voice activation command can be set by the user, like, hey Petra) and it will listen for the command or query from the user.

(b) According to the user’s question, the assistant will search for the answer and provide the best solution possible.

(c) If the assistant is unable to listen, it will again ask the user for some input. This time of asking again can be modified by the user. According to the user's preferences, the assistant can have a male or female voice.

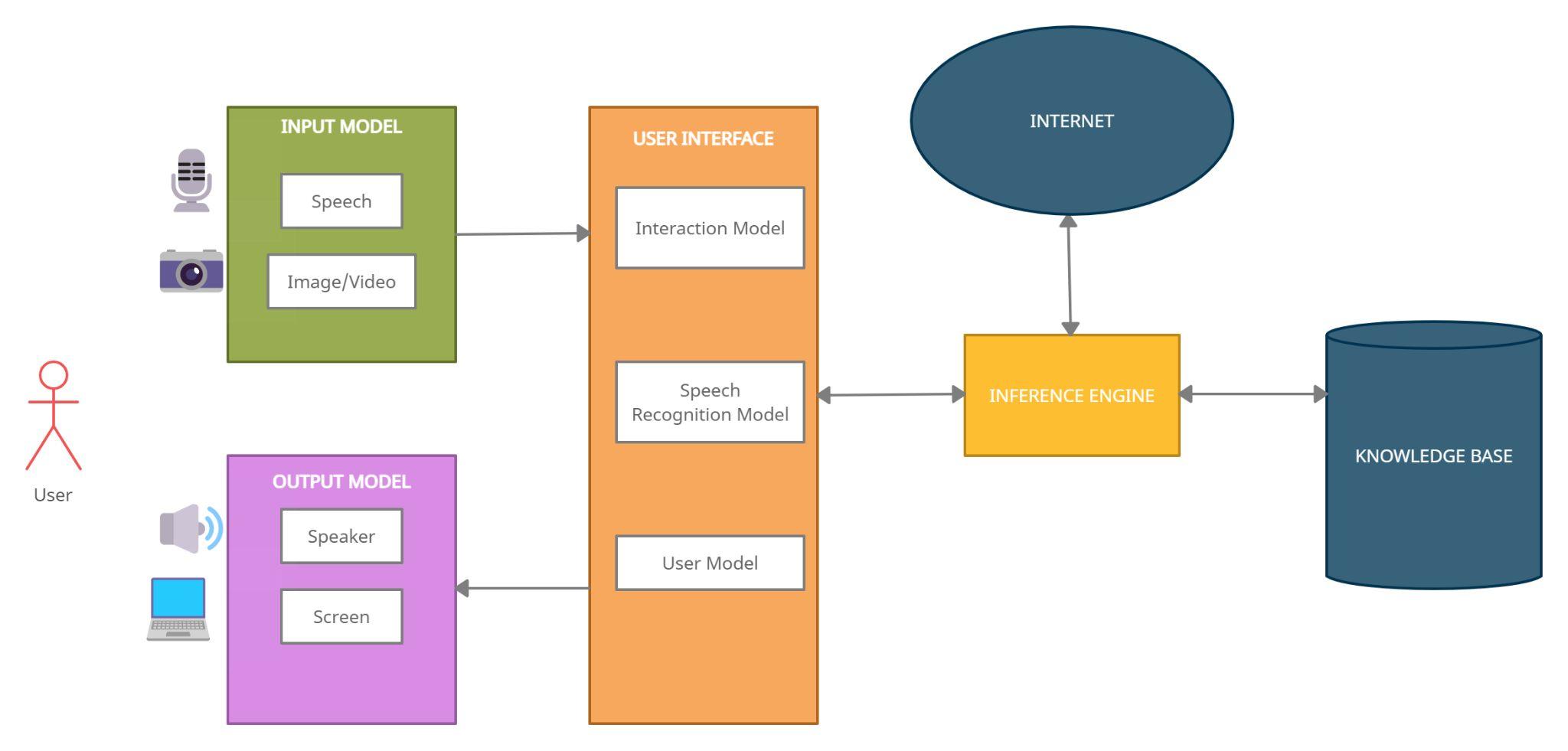
(d) If the assistant is unable to extract information from the user's input, it will ask the user to repeat the process until the desired number of times has been reached.

(e) Finally, if the assistant doesn’t get any input, it will shut itself down. Or if the user is done with it, they can close the assistant just by saying any stop word of their choice (quit, goodbye, exit).

**CHAPTER 4**

**SYSTEM DESIGN**

**4.1 SYSTEM STRUCTURE**

****

**Figure 4.1:** Structure.

**4.1.1 KNOWLEDGE BASE**

There are two knowledge bases. The first is online and the second is a local knowledge base which includes all data and facts based on each model, such as speech recognition knowledge bases, dictionary and spoken dialog knowledge base, user's information and the setting system etc.

**4.1.2 SPEECH RECOGNITION MODEL**

The speech recognition model will work in real-time with the microphone and the input model to recognize the utterances that a user speaks into a microphone and then convert it to text; then it understands the problem and returns the result in both text and voice forms.

**4.1.3 INTERACTION MODEL**

This is the main model that will be used to provide interaction between user and the system by receiving the data from the input model and analyzing it, then returning results.

**4.1.4 INFERENCE ENGINE**

The inference engine works together with the Interaction Model in the chain of conditions and derivations and finally deduces the outcome by analyzing all the facts and rules, then sorting them before concluding to a solution.

**4.1.5 USER MODEL**

This model has all information about all the users that will use the system. It can include personal information such as users' names and ages, their interests, their skills and knowledge, their goals and plans, their preferences and their dislikes or data about their behavior and their interactions with the system. The system will hold only the data that the user wants it to. All information will be collected by asking the user some questions then storing all answers in the Knowledge Base.

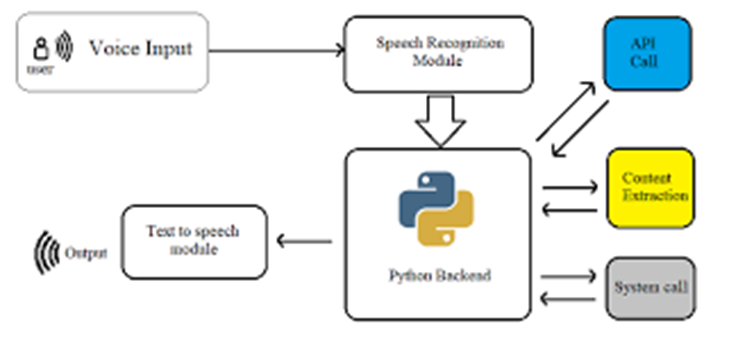
**4.1.6 INPUT MODEL**

This model will organize the work of all input devices that the system uses to collect the different data from microphones of all the devices used, camera (if any query is image/video). Also, this model includes intelligence algorithms to organize the input information before sending the data to the Interaction Model.

**4.1.7 OUTPUT MODEL**

This model will receive the final decision from the Interaction Model, then it will show the result through text, voice, image/video (if applicable).

**4.2 WORKING PRINCIPLES USED**

****

**Figure 4.2:** System Architecture.

**4.2.1 Speech Recognition**

The programme converts speech input to text using python’s speech recognition module. With voice input, users can get texts from specialized corpora arranged on the information center's computer network server, which are briefly kept in the system before being forwarded for speech recognition. The equivalent text is then received and fed to the central processor.

**4.2.2 Python Backend**

The Python backend receives the voice recognition module's output and determines if the instruction or voice output is an API Request and Context Extraction.

**4.2.3 RPA**

Robotic Process Automation (RPA) is a software technology that allows digital operations like data entering, word processing, and chatting to be automated and hence is very helpful for us as it lets us get rid of a lot of manual tasks.

**4.2.4 Cognitive Automation**

It is the application of machine learning that is used to automate processes that would otherwise require physical labor. This with RPA lets us tackle any complex or complicated query. When used with data science techniques it also helps us overcome the accent problem.

**4.2.5 NLP**

Natural language processing is an area of AI that aims to enable computers to interpret text and spoken words in the same manner that humans can and hence it makes our assistant feel natural and not a robot, like google’s assistant.

**4.2.6 API Calls**

A software interface that allows two programmes to interact with one another is known as an API (application programming interface). To put it another way, an API is a messenger that transmits your request to the provider and then returns the response. This lets us scrap the web and use the internet according to the user’s queries or commands.

**4.2.7 Content Extraction**

Context extraction (CE) is the process of obtaining structured data from machine-readable materials that are unstructured or semi-structured. In most situations, this activity entails employing natural language processing to process human language documents (NLP). Test results for context extraction may be seen in recent operations in multimedia document processing, such as automated annotation and content extraction from images, audio and videos.

**4.2.8 Neural Networking**

The assistant may learn about users and become more customized as a result of neural networking with the database maintained, eventually carrying out activities without the user having to ask.

**4.2.9 Machine Learning**

ML aids the assistant in doing many jobs and answering tough queries. VPA also uses machine learning to acquire insight into a person's preferences based on prior decisions and data.

**4.2.10 Text-to-Speech Module**

The capacity of computers to read text aloud is referred to as text-to-speech (TTS). Written text is converted to a phonemic representation, which is subsequently converted to waveforms that may be generated as sound by a TTS Engine. Third-party publishers offer TTS engines in a variety of languages, dialects, and specialist vocabularies.

**4.2.11 Chrome Driver**

It is an open-source tool for automated testing across many browsers. It provides capabilities for navigating to web pages, user input, JavaScript execution, and more.

**4.2.12 Selenium Web Automation**

Selenium is an open-source umbrella project for a range of tools and libraries aimed at supporting browser automation. It provides a playback tool for authoring functional tests without the need to learn a test scripting language.

**4.2.13 SQLite**

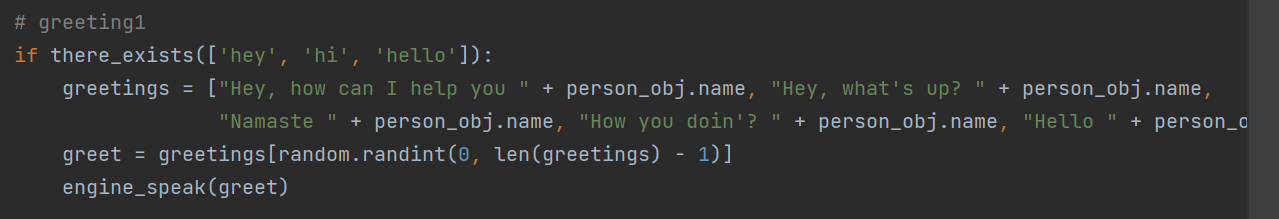
SQLite is a database engine written in the C language. It is not a standalone app; rather, it is a library that software developers embed in their apps. As such, it belongs to the family of embedded databases. We have used SQLite to maintain the database of every user and the knowledge base.

**CHAPTER 5**

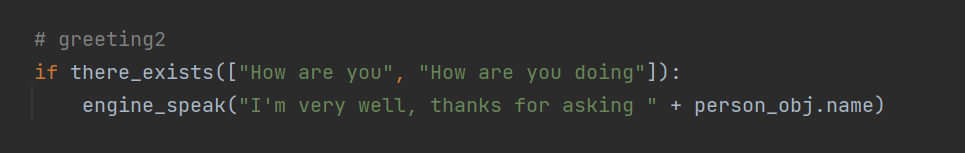
**IMPLEMENTATION**

The assistant, on starting, will initially wait for the input to be given from the user. If the user gives input command, via voice, the assistant will capture it, and search for the keyword present in the input command. If the assistant is able to find a keyword, then it will perform the task accordingly, and return the output back to the user, in voice and in text. If not, the assistant will again start waiting for the user to give input. Each of these functionalities have their own importance in the whole system working. Below are some of the important tasks that the user will require on a daily basis, that our assistant can perform.

**5.1 GREETING**

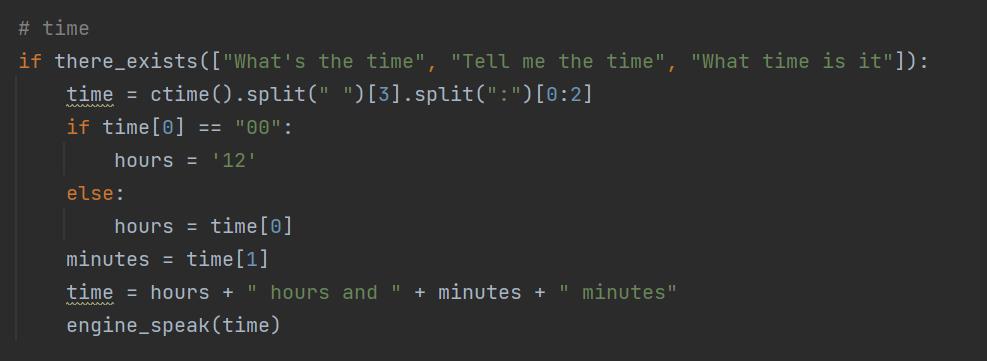
**Figure 5.1:** Greeting 1.

If the user greets the assistant by using words such as, “hey”, “hi” or “hello”, the assistant greets the user back by saying “Hey, how can I help you?”, “Hey, what’s up?” etc. This makes the user feel at ease and it also lets the user feel like they are talking to an assistant not a robot.

**Figure 5.2:** Greeting 2.

If the user asks for the well-being of the assistant, the assistant replies in a friendly manner that “I’m very well, thanks for asking”. Greeting is one of the basic functions of communication and triggers positive conversations. It helps us connect to people at a more personal level. Greetings can be so powerful that it can even turn your frown into a smile and drastically put you in a good mood, hence, it is important for the assistant to greet the users everytime they use it.

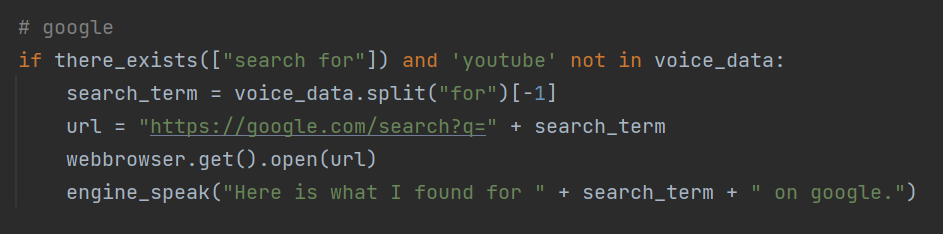
**5.2 TIME**

****

**Figure 5.3:** Time.

If the user asks the assistant for the time, by saying basic phrases like, “What’s the time?”, “What time is it?” etc. The assistant quickly replies with the current time. This helps the user to keep track of the time and helps the time management. Effective time management increases your focus and improves your productivity. Greater focus allows you to capture bigger opportunities. It also allows you to spend more time on the projects, goals, and people that matter. Time management is important in helping you achieve greater focus and prioritization.

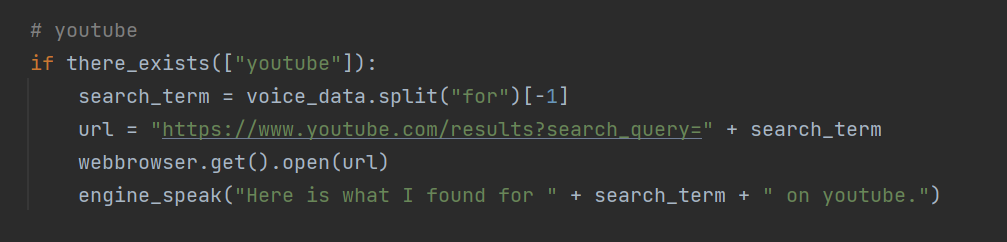
**5.3 GOOGLE**

****

**Figure 5.4:** Google.

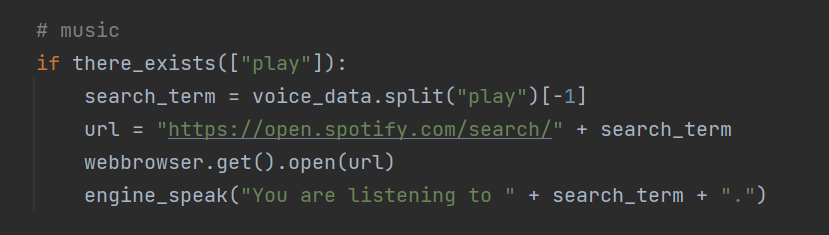
If the user wants to search for anything then our assistant provides the answer using google search engine. It traces the key phrase “search for” and intelligently understands the query and quickly opens the web browser and gives the result. Today Google is no less than an encyclopedia that offers you information or knowledge on everything under the sky, we all know the importance of google as every one of us at least use google once a day.

**5.4 YOUTUBE**

**Figure 5.5:** YouTube.

Suppose if the user is bored or if the user wants to watch some video on the internet. There’s only one place to go to and that’s youtube. YouTube is the 2nd largest search engine next to Google. People upload more than 100 hours of video per minute to YouTube. It's one of the best ways to communicate to a wide audience, whether you're promoting programs or providing information to students. Just by saying youtube followed by the name of the video, our assistant lets the user search for anything on it.

**5.5 MUSIC**

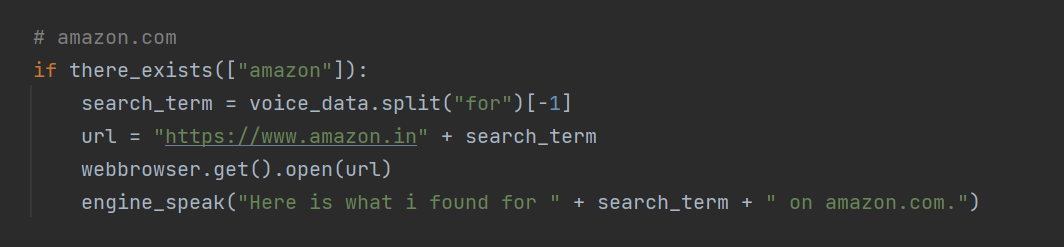
****

**Figure 5.6:** Music on Spotify.

If the user wants to listen to some music, they just have to say play followed by the name of the song and our assistant will play it on Spotify. Spotify is one of the largest music streaming service providers, with over 422 million monthly active users, including 182 million paying subscribers.

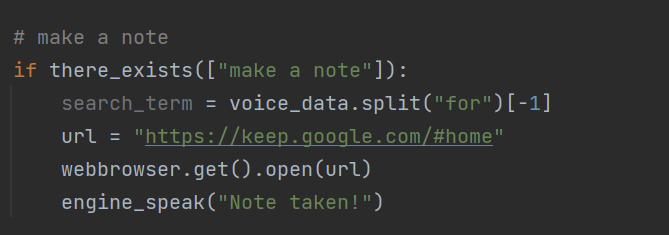
**5.6 AMAZON**

If the user wants to shop something or wants to look for something on the internet, there’s basically only one place to go to and that’s Amazon.com. The assistant will look for the keyword “amazon” and identify the search term by itself from the input given and will show you the results on Amazon.in. Amazon has been referred to as "one of the most influential economic and cultural forces in the world", and is one of the world's most valuable brands.



**Figure 5.7:** Amazon.

**5.7 MAKE A NOTE**



**Figure 5.8:** Note taking on Google Keep.

If the user is in a hurry or they want to take a note of something important to look at in the future, they can do that by just saying “make a note”, and the assistant will make the note on Google Keep. Google Keep is a note-taking service included as part of the free, web-based Google Docs Editors suite offered by Google.

**5.8 SOCIAL MEDIA**

We all know the power of social media in today’s life, and almost everyone spends most of their free time on social media. Our assistant lets you enjoy the services of Instagram and Twitter just by saying “open instagram” or “open twitter”. Instagram allows users to upload media that can be edited with filters and organized by hashtags and geographical tagging. Twitter lets users post, like, and retweet tweets.



**Figure 5.9:** Instagram and Twitter.

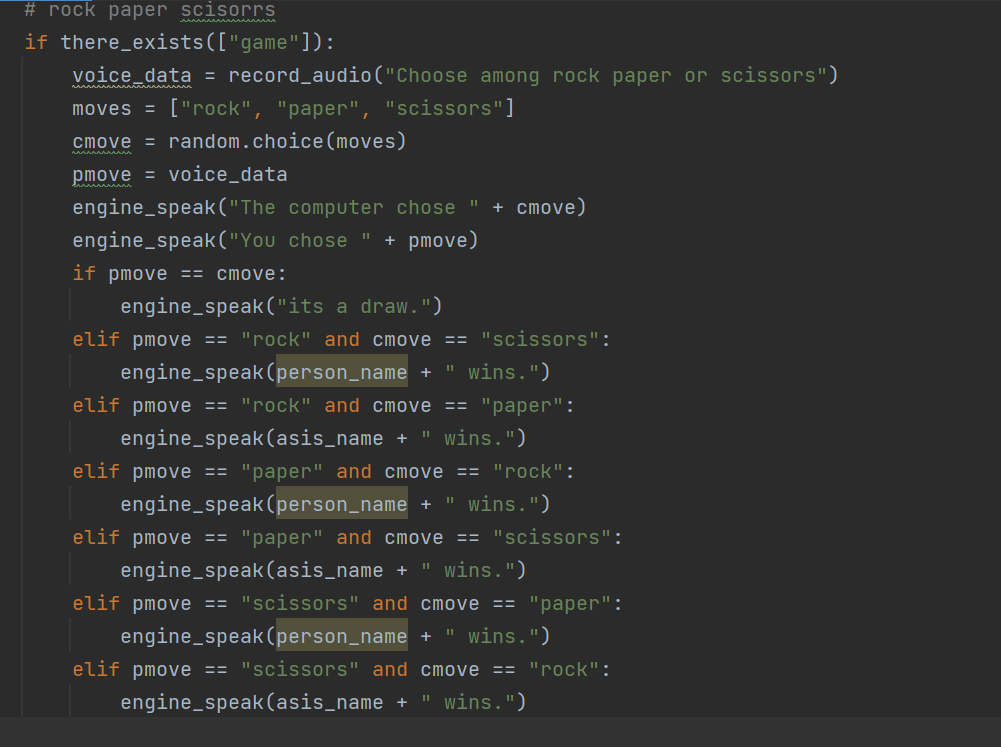
**5.9 WEATHER**



**Figure 5.10:** Weather.

When the user says words like “weather”, “how’s the weather?” etc. The assistant searches for the current temperature on google. It also shows precipitation, humidity and wind speed. Moreover, it also gives the weather report for the upcoming days. This lets the user know of any possible weather warnings. Weather warnings are important forecasts because they are used to protect life and property. Forecasts based on temperature and precipitation are important to agriculture, and therefore to traders within commodity markets.

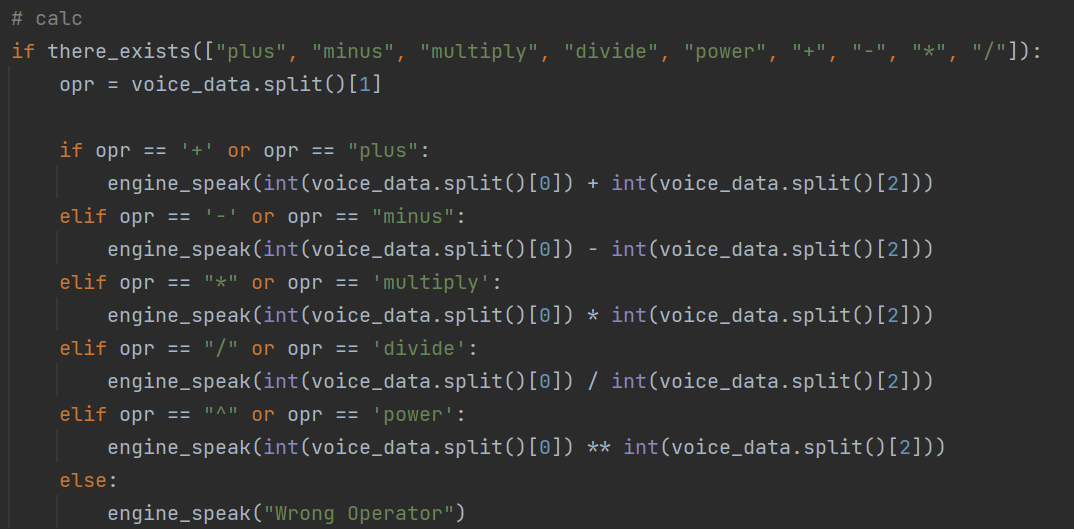
**5.10 GAME**



**Figure 5.11:** Rock, Paper and Scissors.

The assistant also offers a small and fun game of rock, paper and scissors for the user if they are bored and want to have some fun or have a small break from the work. User just has to mention “game” and then the assistant asks the user to pick one from rock, paper or scissors. Users play against the assistant, hence showing again that it's not just a robot you are interacting with. Rock Paper Scissors is a hand game originating from China, usually played between two people, in which each player simultaneously forms one of three shapes with an outstretched hand. These shapes are "rock", "paper", and "scissors".

**5.11 CALCULATOR**

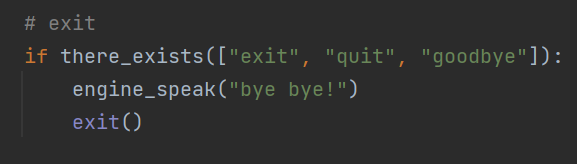


**Figure 5.12:** Calculator.

Our assistant also has an in-built calculator to perform any quick mathematical operation that the user might have. Calculators help solve more problems in less time. Calculators allow students to work more quickly, which means they can solve more problems in a given time.

**5.12 EXIT**

Just by mentioning any of the three stop words, “exit”, “quit” or “goodbye”, the user can shut the assistant off. In reply the assistant nicely answers “bye bye!” and shuts itself down.



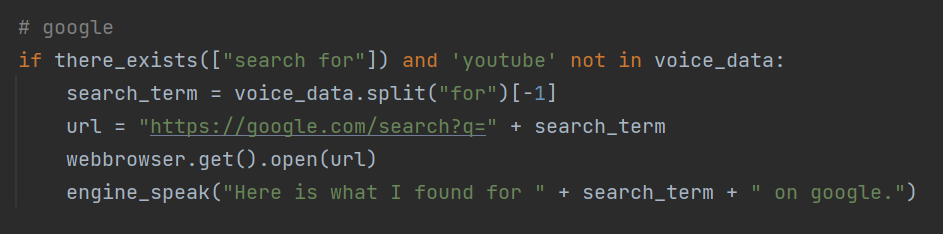
**Figure 5.13:** Exit.

**CHAPTER 6**

**RESULTS AND DISCUSSION**

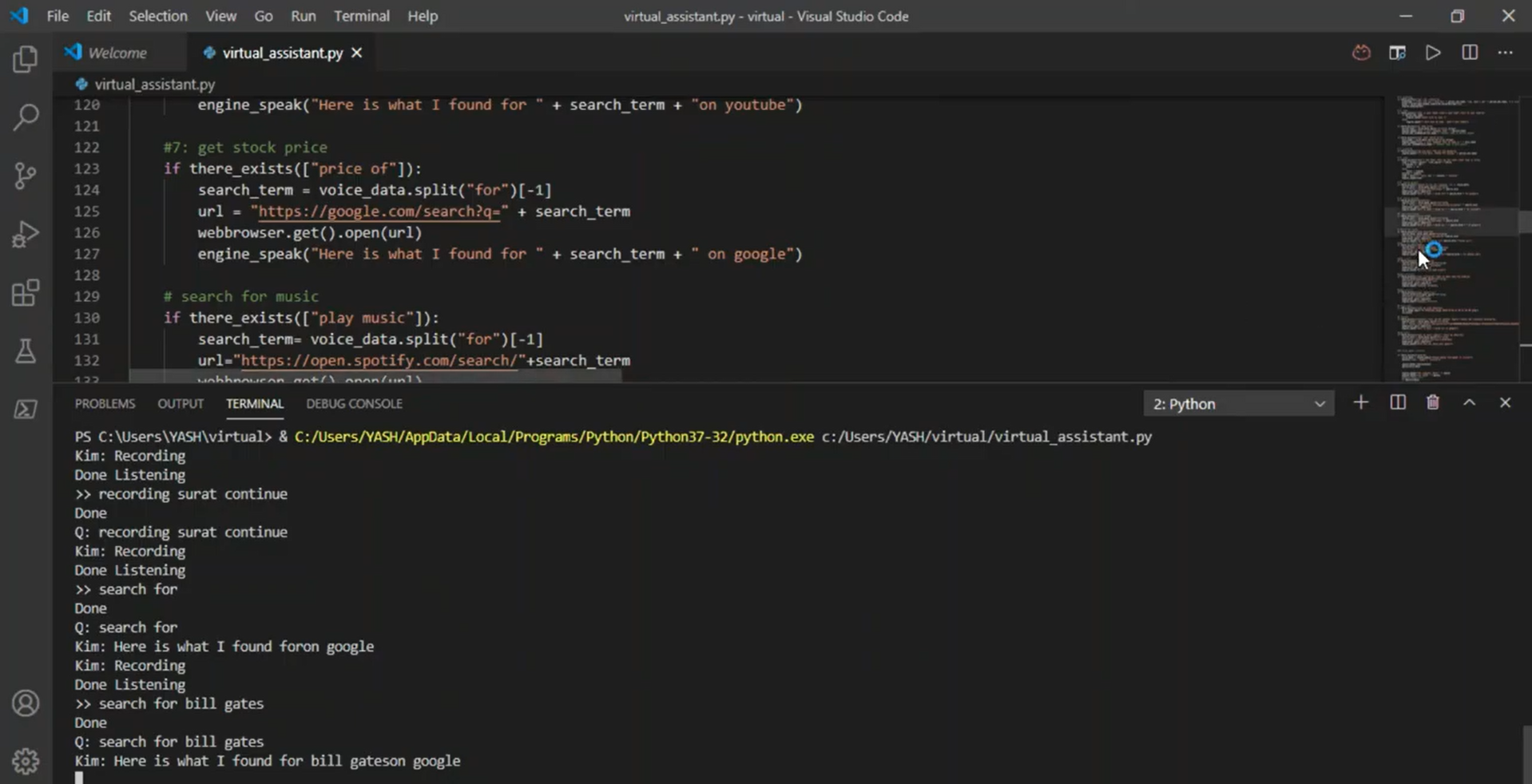
This section provides a brief description of our result on the basis of the comparison and analysis of our proposed work. We have employed this idea by the means of RPA, AI, Python, and ML. Our assistant currently works on the desktop system. Anyone with Windows 7 or above can use it. It has been tested on the different versions of Microsoft Windows, such as Windows 7, Windows 8, Windows 8.1, Windows 10, and Windows 11. And on each of these operating systems, our assistant has worked efficiently. Our assistant is very reliable as it performs each of the operations listed out in Chapter 5 on each of the operating systems mentioned above. This assistant completes each and every task that has been mentioned, so its validity is justifiable.

Here’s an example of how our assistant works.

****

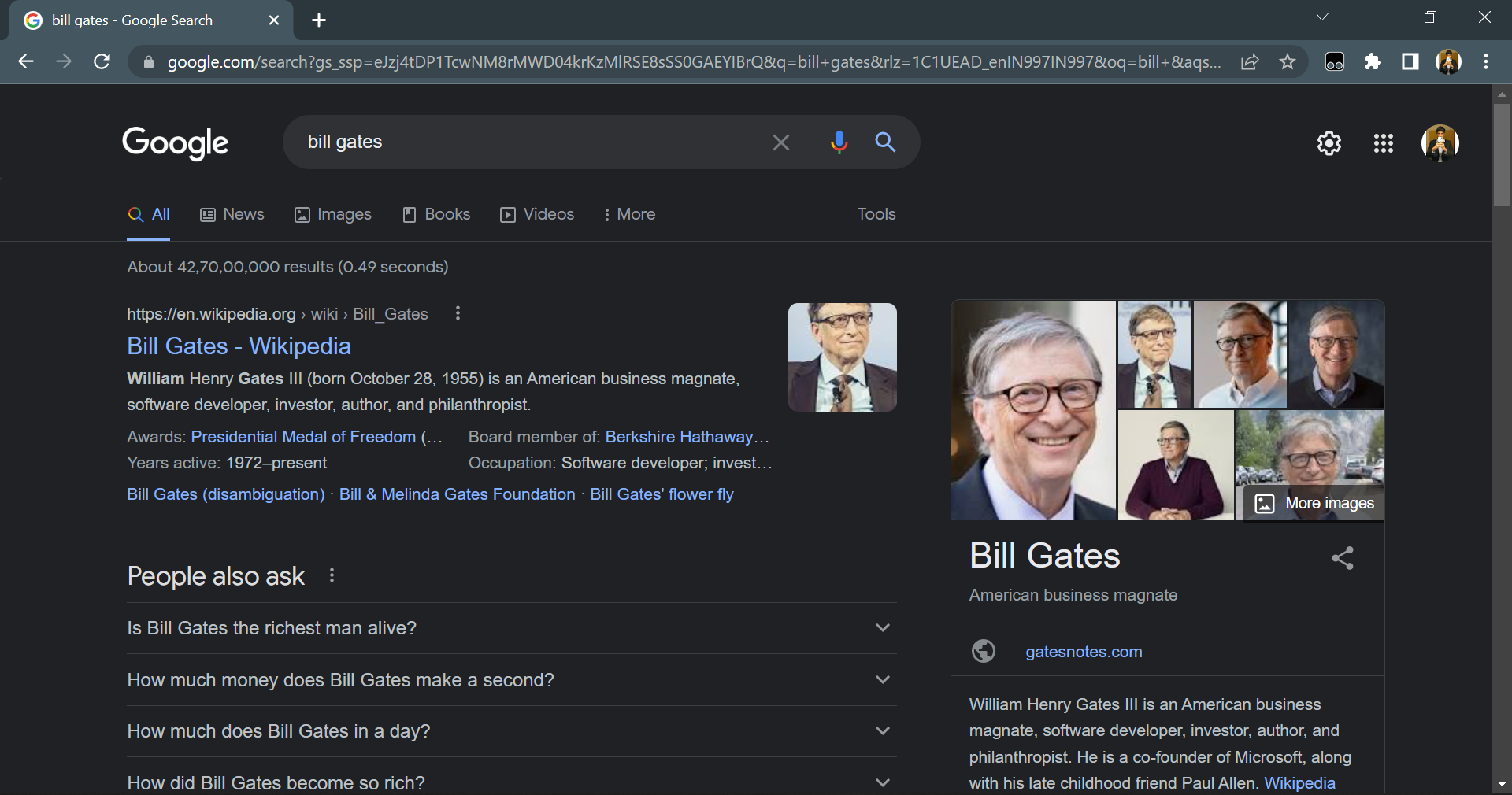
**Figure 6.1:** Google.

This is the code that is written for the assistant to search on Google. When the user wants to search for something on Google, they usually say "search for" followed by whatever they want to search for. So our intelligent assistant identifies the term "search for" and understands that the user wants to access Google. So, the VPA splits the query and searches for whatever the search term is given by the user.



**Figure 6.2:** Assistant Working.

The assistant first starts recording, and after the user is done speaking, it lets the user know that it is done listening. In the meantime, it identifies the query and produces the desired result. Note that the assistant’s default name is set to "Petra" and the user could change it anytime. Here it shows Kim because during the testing we might’ve checked this name-changing functionality.



**Figure 6.3:** Result.

In response to the above query, the assistant opens the default browser and shows the result on Google.

**TESTING**

Testing is the process of executing a program with the aim of finding errors. To make our software perform well it should be error-free. If testing is done successfully, it will remove all the errors from the software.

1. **UNIT TESTING -** It focuses on the smallest unit of software design. In this, we test an individual unit or group of interrelated units. It is often done by the programmer by using sample input and observing its corresponding outputs. We tested each and every snippet of our loop, checking every functionality that has been provided one-by-one to ensure there were no errors. We checked whether every loop, method, and function was working fine or not. We also made sure that there wasn’t any incorrect initialization or indentation present.
2. **BLACK BOX TESTING -** It is used for validation. In this we ignored all the internal working mechanisms and focused on what the output was. So we checked each and every functionality of our VPA and found out that nothing was wrong.
3. **WHITE BOX TESTING -** It is used for verification. In this we focus on internal mechanisms, i.e., how is the output achieved? We have used normal web scraping for most of our functionalities, so it does not violate anything.
4. **ALPHA TESTING -** This is a type of validation testing. It is a type of acceptance testing that is done before the product is released to customers. We ensured that there wasn’t anything wrong before handing our assistant over for beta testing.
5. **BETA TESTING -** The beta test is conducted at one or more customer sites by the end-user of the software. We tested our assistant out with our classmates and found no problems or complaints at all.
6. **SYSTEM TESTING -** This software is tested so that it works fine for different operating systems. We have mentioned above that it works fine on any of the Microsoft Windows versions after Windows 7.

Through this voice assistant, we have automated various services using a single line command that can be given verbally. It eases most of the tasks of the user, like searching the web, retrieving weather forecast details, vocabulary help etc. It not only works on human commands but also gives responses to the user based on the query being asked or the words spoken by the user, such as opening tasks and operations. It is a way of greeting the user the way the user feels more comfortable and feels free to interact with the voice assistant. The application should also eliminate any kind of unnecessary manual work required in the user's life when performing every task. The entire system works on verbal input rather than the next one.



**Figure 6.4:** Voice enabled VPA.

The Assistant works properly to perform any tasks given by the user. It also increases the interaction between users and computers. This system also remembers each and every user because of the knowledge base we have created. Just by asking some generic personal questions, it truly becomes a virtual *personal* assistant.

**CHAPTER 7**

**CONCLUSION AND SUMMARY**

In this project, “Virtual Personal Assistant Using RPA, AI, and Python”, we designed and implemented digital assistance. The project is developed with open-source software components that are supported by the PyCharm community and can quickly accept any modifications. Because this project is modular, it is more adaptable and easier to add new features without disrupting existing system functionality. It not only responds to human instructions but also delivers replies depending on the inquiry or words said by the user, such as opening tasks and operations. It greets the user in a way that makes them feel more at ease and allows them to communicate freely with the voice assistant. The application should also reduce any extra manual labor that is required in the user's day-to-day tasks. The whole system is based on verbal input.

This virtual assistant can improve efficiency and offer support to both employees and customers of the organization. It allows them to offer more services by taking over more of the routine tasks. Hence, employees can spend more time on other tasks. It not only allows them to offer more and better services, but it also allows the organization to save money. This virtual assistant can increase the safety factor in automobile engineering as it can eliminate the physical interaction of the driver with the machine. They can easily communicate with the VPA through their voice, which can handle almost any query or command.

**CHAPTER 8**

**FUTURE SCOPE**

Virtual assistants are becoming a popular and useful technology, with a variety of advantages, contributing to the automation of tasks and providing support in time-management, accessing information, communication facilitation, etc. The technology is still in its early stages. The Virtual Assistant presented has few new functionalities compared to other assistants, but an additional and forward-looking feature may be embedded as a future work of this project. This assistant's future upgrades will combine IoT to operate nearby objects, similar to how Amazon's Alexa works. We will also try to get a version of our assistant which works on the Linux systems also, as there doesn’t exist any virtual assistant for these operating systems. It has been predicted that by 2025, 50% of knowledge workers will use a virtual assistant on a daily basis. And by 2023, 25% of employee interactions will be voice-based communications. We can also add some interesting and new functionalities such as graphic designing and blog writing which will help the users even more. We would also like to overcome a couple of hurdles in the future. First, the security of the voice transmissions needs to be addressed, so that businesses will feel comfortable accessing confidential data through voice command. Second, we need to increase the comfort level of these assistants even more, so they can be looked at as useful tools and not a time passing gimmick.

This is only the beginning; as AI and machine learning advance, our virtual assistant will get smarter and provide new possibilities.

**APPENDIX**

**A.1 SNIPPET OF CODE OF VIRTUAL ASSISTANT**

#gmail

if there\_exists(["open my mail","open gmail","check my email"]):

try:

search\_term = voice\_data.split("mail")[-1]

except:

search\_term = voice\_data.split("gmail")[-1]

url="https://mail.google.com/mail/u/0/#inbox"

webbrowser.get().open(url)

engine\_speak("opening gmail...")

#rock paper scissors

if there\_exists(["game"]):

voice\_data = record\_audio("Choose among rock paper or scissors")

moves=["rock", "paper", "scissors"]

cmove=random.choice(moves)

pmove=voice\_data

engine\_speak("The computer chose " + cmove)

engine\_speak("You chose " + pmove)

if pmove==cmove:

engine\_speak("its a draw.")

elif pmove== "rock" and cmove== "scissors":

engine\_speak(person\_name + " wins.")

elif pmove== "rock" and cmove== "paper":

engine\_speak(asis\_name + " wins.")

elif pmove== "paper" and cmove== "rock":

engine\_speak(person\_name + " wins.")

elif pmove== "paper" and cmove== "scissors":

engine\_speak(asis\_name + " wins.")

elif pmove== "scissors" and cmove== "paper":

engine\_speak(person\_name + " wins.")

elif pmove== "scissors" and cmove== "rock":

engine\_speak(asis\_name + " wins.")

**A.2 STATEMENTS AND DECLARATIONS**

**A.2.1 FUNDING**

We wish to confirm that there has been no significant financial support for this work that could have influenced its outcome.

**A.2.2 CONFLICT OF INTEREST**

We wish to confirm that there are no known conflicts of interest associated with this publication.

**A.2.3 ETHICS APPROVAL**

We further confirm that the work covered in this manuscript has not involved either experimental animals or human patients. Ethics approval not applicable to this project.

**A.2.4 CONSENT TO PARTICIPATE**

We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed.

**A.2.5 CONSENT FOR PUBLICATION**

We confirm that there’s no impediments for publishing the submitted manuscript.

**A.2.6 CODE AVAILABILITY**

We confirm that corresponding codes and the dataset generated for the work done in this manuscript are available from the corresponding author on reasonable request.

**BIBLIOGRAPHY**

[1] A. Sudhakar Reddy M., Vyshnavi, C. Raju Kumar, and Saumya, "VIRTUAL ASSISTANT USING ARTIFICIAL INTELLIGENCE ", International Journal of Emerging Technologies and Innovative Research (www.jetir.org), ISSN:2349-5162, Vol.7, Issue 3, page no. 1116-1119, March-2020.

[2] George Terzopoulos, Maya Satratzemi, “Voice Assistants and Smart Speakers in Everyday Life and in Education”, Informatics in Education, v19 n3 p473-490 2020.

[3] Regina Gubareva, Rui Pedro Lopes, “Virtual Assistants for Learning: A Systematic Literature Review”, CSEDU20-RP-93, May-2020.

[4] Emad S. Othman. “Voice Controlled Personal Assistant Using Raspberry Pi”. International Journal of Scientific and Engineering Research Volume 8, Issue 11, November-2017.

[5] V. Radha and C. Vimala, “A review on speech recognition challenges and approaches” doaj.org, vol. 2, no. 1, pp. 1–7, 2012.

[6] Bassam A. & Raja N. (2010) “Arabic speech recognition using Hidden Markov Model Toolkit (HTK)”. 2. 557 - 562. 10.1109/ITSIM.2010.5561391.

[7] Elshafei, M., Virtual personal assistant (VPA) for mobile users. Mitel Networks (2000–2002).

[8] B. S. Atal and L. R. Rabiner, “A pattern recognition approach to voiced unvoiced-silence classification with applications to speech recognition” Acoustics, Speech and Signal Processing, IEEE Transactions on, vol. 24, no. 3, pp. 201–212, 1976.

[9] Jatu Naazneen, Abdul Gaffar, Abhijit Palse, “Virtual Assistant using Python”, DTSS COLLEGE OF COMMERCE, 2018-2019.