# A PERSONAL VOICE-ACTIVATED DESKTOP ASSISTANT

## A Project Work Synopsis

*Submitted in the partial fulfilment for the award of the degree of*

## BACHELOR OF ENGINEERING

**IN**

**ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING**

**Submitted by:**

**MAYANK GUPTA (20BCS6892)**

**AAYUSH ADHIKARI (21BCS8881)**

**MITALI GUPTA (20BCS6890)**

**KARAN JAIN (21BCS8804)**

**Under the super vision of:**

**AMIT GARG**



**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**APEX INSTITUTE OF TECHNOLOGY**

**CHANDIGARH UNIVERSITY, GHARUAN, MOHALI - 140413,**

**PUNJAB**

#### MONTH & YEAR

#### FEBRUARY, 2022

**ABSTRACT**

Everything in the twenty-first century specifically is automated, including everyday stuff pretty such as bus doors, air conditioning systems, and turning everything on with a kind of single click, to name a sort of few in a particularly big way.

In this fast-paced world, the very current study proposes a newer concept of a voice-controlled gadget that recognizes one's speech, processes the request, and details particularly other related information in a sort of major way. We need to basically build gadgets with generally built-in voice recognition, as well as a facial recognition system, that can really detect a person's speech even in a busy situation, which kind of is quite significant.

We basically thought it would generally be amusing to specifically create a personal assistant in Python, demonstrating how everything in the twenty-first century generally is automated, including everyday stuff generally such as bus doors, air conditioning systems, and turning everything on with a fairly single click, to name a definitely few in an actually major way.

The gadget will essentially listen to noises through its microphone, process the human's inquiry, and essentially react with the relevant information, which basically is fairly significant.

If you for all intents and purposes ask the device to generally play a video on basically your computer, it will, for example, access YouTube and particularly play the video you choose, demonstrating that we need to literally build gadgets with kind of built-in voice recognition, as well as a facial recognition system, that can mostly detect a person's speech even in a busy situation, which definitely is quite significant.

Speech recognition may actually be done with the Python Speech Recognition module, which for the most part is quite significant. We essentially utilize the Google Speech API because of it’s for all intents and purposes great quality, demonstrating that we need to all intents and purposes build gadgets with definitely built-in voice recognition, as well as a facial recognition system, that can literally detect a person's speech even in a busy situation in a subtle way.

**Table of Contents**

Title Page i

Abstract ii

* INTRODUCTION
  + Problem Definition
  + Requirement Specification
  + Hardware and Software Requirements
  + Modules Imported
* LITERATURE REVIEW
  + Siri (Apple)
  + Alexa (Amazon)
  + Bixby (Samsung)
* PROBLEM FORMULATION
* OBJECTIVE
* METHODOLOGY
  + System Architecture
  + Methodology for voice email capability
* EXPERIMENTAL SETUP
* Feasibility Study
* RESULTS
* CONCLUSION AND FUTURE SCOPE
* REFERENCES

#### INTRODUCTION

Almost all duties are now digitalized in today's world. We have a Smartphone in our hands, and it's like having the entire world at our fingertips. We don't even use our fingers anymore.

We just speak of the task and it is done. There are procedures in place where we can text Dad and say, "I'll be late today." And the text is sent. A Virtual Assistant's job is to do just that. It also helps automate search, discovery, and online purchase processes by supporting specific tasks such as booking a trip or locating the cheapest book online from several eCommerce sites and then giving an interface to place an order.

Virtual assistants are software programs that assist you with day-to-day duties such as weather forecasting, setting reminders, and preparing shopping lists, among other things. They can accept text (online chatbots) or voice commands. To activate the listener, voice-based intelligent assistants require an invoking phrase or wake word, followed by the command. There is a plethora of virtual assistants available, including Apple's Siri, Amazon's Alexa, and Microsoft's Cortana

This system is intended to be used on desktop computers. Personal assistant software helps users be more productive by handling their everyday chores and providing them with information from web sources. And it's done in a matter of seconds.

Voice searches have surpassed text searches in popularity. Web searches conducted via mobile devices have just eclipsed those carried out using a computer and the researchers are already estimating that 50 percent of queries will be via voice by 2020. Virtual assistants are proving to be more intelligent than ever. Allow your intelligent assistant to do the heavy lifting for you when it comes to email.

Detect intent, extract key data, automate processes, and provide customized solutions.

This project was started on the assumption that there is enough freely available data and information on the internet to develop a virtual assistant capable of making intelligent decisions for ordinary user tasks.

* **PROBLEM DEFINITION**:

In most cases, a user must manually manage many sets of applications in order to execute a single operation. A user planning a trip, for example, should look up airport codes for neighbouring airports and then search travel sites for tickets between airport combinations to get to their destination. A system that can effortlessly manage chores is required.

We already have a number of virtual assistants on staff. However, we hardly ever utilise it. A large number of people struggle with speech recognition. These systems are capable of understanding English phrases, but they are unable to distinguish our dialect. Our pronunciation differs significantly from theirs.

They are also more user-friendly on mobile devices than desktop systems. A virtual assistant who can understand English with an Indian accent and work on a desktop system is required.

When a virtual assistant is unable to effectively answer inquiries, it is due to a lack of context or an understanding of the question's intent. It can only answer relevant questions through thorough optimization integrating both humans and machine learning. Continually maintaining solid quality control procedures will also aid in reducing the possibility of the virtual assistant picking up unwanted negative habits. In order for them to function well, they must be given a tremendous volume of data.

Virtual assistant 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 the it should be able to consider user preferences, other active tasks, priorities in order to recommend a particular plan.

* **REQUIREMENT SPECIFICATION:**

Personal assistant software must operate as a portal into the digital world, understanding user requests or commands and turning them into actions or recommendations based on the agent's knowledge of the world.

Our virtual assistant focuses on removing the need for the user to type text input and instead relying on speech as the primary mode of user input. The agent then records the input after applying voice recognition techniques to it. It then uses this information to call one of the personal information management programmes, such as a task list or calendar, to create a new entry or to conduct a search on Google, Bing, or Yahoo, among others. The emphasis is on gathering user input via voice, recognising it, and then completing tasks if the agent understands them. Software interprets this data in normal language, making it easier for the user to specify what he or she wants done.

Voice recognition software enables hands free use of the applications, lets users to query or command the agent through voice interface. This helps users to have access to the agent while performing other tasks and thus enhances value of the system itself. Our Assistant will also have 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.

Virtual assistants must provide a wide variety of services. These include:

 • Providing information such as weather, facts from e.g., Wikipedia etc.

• Set an alarm or make to-do lists and shopping lists.

 • Remind you of birthdays and meetings.

 • Play music from streaming services such as Saavan and Gaana.

 • Play videos, TV shows or movies on televisions, streaming from e.g., Netflix or Hotstar.

 • Book tickets for shows, travel and movies.

* **HARDWARE AND SOFTWARE REQUIREMENTS:**

**HARDWARE:**

* A good computer or laptop, with at least i3 processor for running IDEs.
* RAM of at least 1gb.
* Microphone.

**SOFTWARE:**

* Window 7 or later.
* Python 3 language.
* IDE like PyCharm, Spyder or Visual Studio Code.
* **Modules Imported:**
* **WolframAlpha**- It is used to compute expert-level answers of any command using Wolfram’s algorithms, knowledgebase and AI technology.
* **JSON**- JavaScript Object Notation. JSON is a lightweight format for storing and transporting data. JSON is used when data is sent from a server to a web page. JSON is "self-describing" and easy to understand
* **Speech recognition**- Speech recognition means that when humans are speaking, a machine understands it. In our project we are using Google Speech API in Python to make software which is used to run machines on command. We need to install the PyAudio python package for recognize the voice commands. PyAudio is installed using pip install PyAudio command.
* **gTTS**- Google’s text-to-speech packages converts your audio questions command to text. The response from the look-up function that you write for fetching answer to the question or command is converted in an audio form by gTTS. This package interface with Google Translate’s API.
* **Datetime**- Datetime package is used to showing Date and Time. This datetime module comes with built in Python.
* **Wikipedia**- We all know Wikipedia is a great and huge source of knowledge just like GeeksforGeeks or any other sources we have used the Wikipedia module in our project to get more information from Wikipedia or to perform a Wikipedia search. To install this Wikipedia module use pip install Wikipedia.
* **WebBrowser**- To perform Web Search. This module comes built-in with Python.
* **OS**- The OS module in Python provides functions for interacting with the os. OS comes under Python’s standard utility modules. This module provides a way of using operating system dependent functionality.
* **Pyjokes**- Pyjokes is used for collection Jokes over the Internet. Pyjokes is add in our project because it adds jokes in our project. It is very interesting. Pyjokes is the one-line joke which makes our project interesting.
* **PyAudio**- PyAudio is a set of Python bindings for PortAudio, a cross-platform C++ library interfacing with audio drivers.
* **Smtplib**- The simple mail transfer protocol library is a Python library for sending emails using the Simple Mail Transfer Protocol (SMTP). The smtplib is a built-in module in python; do not need to install it. It abstracts all the complexities of SMTP away. It provides a Simple Mail Transfer Protocol (SMTP) client implementation.
* **Requests**- Requests module allows you to send http requests using python. It is used for making GET and POST requests. It abstracts the complexities of making requests behind a beautiful, simple API.

**LITERATURE REVIEW**

A variety of desktop virtual assistants are already available. This section discusses a few instances of existing virtual assistants on the market, as well as the duties they can perform and their limitations.

* **SIRI (Apple)**

SIRI is a voice-activated personal assistant that interacts with the user, detects instructions, and acts on them. It enhances voice recognition over time by learning to adapt to the user's speech. When it can't figure out what the user wants, it attempts to talk to them.

It works with the device's calendar, contacts, and music library applications, as well as the device's GPS and camera. It makes use of spatial, temporal, social, and task-based contexts to tailor the agent's behaviour to the user at any given time.

* **Tasks that are supported**
* Call someone from my contacts list
* Open an app on my iPhone
* Send a text message to someone
* Schedule a meeting at 9 a.m. tomorrow on my calendar Set your alarm for 5 a.m. tomorrow.
* I'd want to play a certain song from my iTunes library.
* Make a new note.
* **Drawback**

SIRI does not have its knowledge store and relies on information recorded in domain and data models to comprehend the world.

* **Alexa (Amazon)**

Alexa is a virtual assistant that you may manage with your voice. She can play music, control your smart home, answer questions, and connect you to your favourite services, keeping you organized, informed, protected, and entertained. She's also your shopper because she's an Amazon product.

Alexa is a cloud-based voice assistant that can be used with a growing variety of smart speakers and other Alexa-enabled devices.

* **Tasks that are supported**
* Reminders
* Email
* Calendar, Google Calendar
* Outlook
* Evernote
* Facebook, LinkedIn
* News Feeds
* **Drawback:**

You can't customize your virtual assistant's name and you're limited to six wake words you can use.

* **Bixby (Samsung)**

Bixby is a virtual assistant that makes it easier to use your phone, giving you more time to focus on what matters the most. Bixby learns, evolves and adapts to what you like to do, working alongside your favourite apps and services to help you get more done.

* **Tasks that are supported**
* Call someone from my contacts list
* Download any app.
* Send a text message to someone
* Schedule a meeting at 9 a.m. tomorrow on my calendar.
* Set your alarm for 5 a.m. tomorrow.
* I'd want to play a certain song
* Write an email.
* Make a new note.
* **Drawbacks:**

Bixby Voice Assistant isn’t as smart as Google Assistant. According to the research, Bixby has security issues, failing to protect your personal information and pictures compared to other personal voice assistant technologies.

**PROBLEM FORMULATION**

In most cases, a user must manually manage many sets of applications to execute a single operation. For example, a person planning a trip should look for airport codes for neighbouring airports and then look for tickets between those airports on travel websites, arrive at the desired location. A system that can effortlessly manage chores is required.

We already have several virtual assistants on staff. However, we hardly ever utilize it. A large number of people struggle with speech recognition. These systems are capable of understanding English phrases, but they are unable to distinguish our dialect.

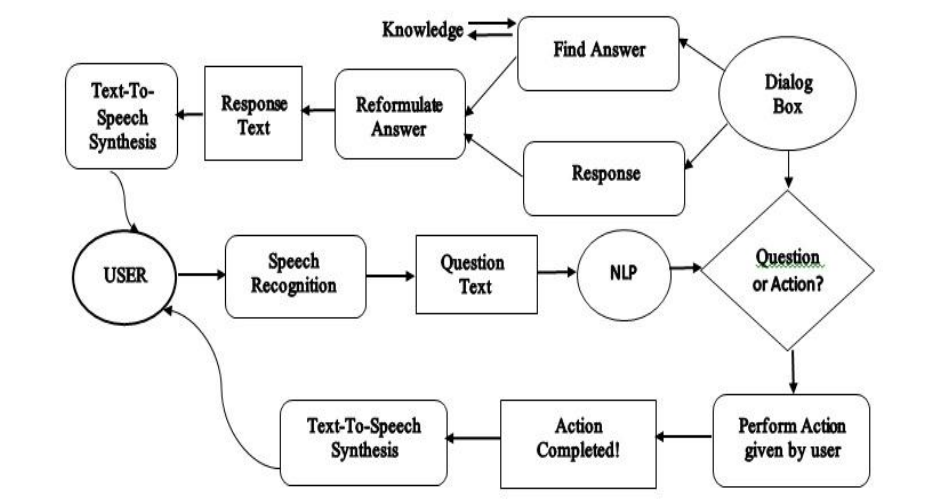
Our pronunciation differs significantly from theirs. They are also more user-friendly on mobile devices than desktop systems. A virtual assistant who can understand English with an Indian accent and work on a desktop system is required.

When a virtual assistant is unable to effectively answer inquiries, it is due to a lack of context or an understanding of the question's intent. It can only answer relevant questions through thorough optimization integrating both humans and machine learning.

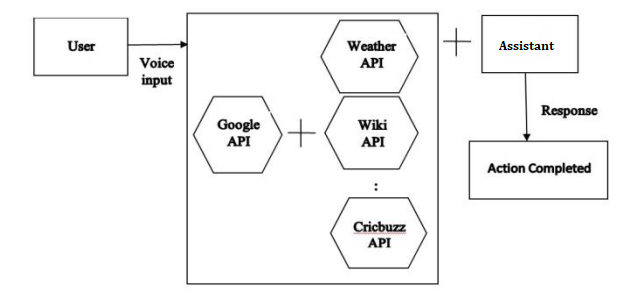
Continually maintaining solid quality control procedures will also aid in reducing the possibility of the virtual assistant picking up unwanted negative habits. For them to function properly, they require a tremendous amount of data to be supplied to them.

Virtual assistants should be able to model complex task dependencies and use these models to suggest user-friendly plans. When a task has several sub-tasks, each of which can have its sub-tasks, it must be tested to determine the best paths.

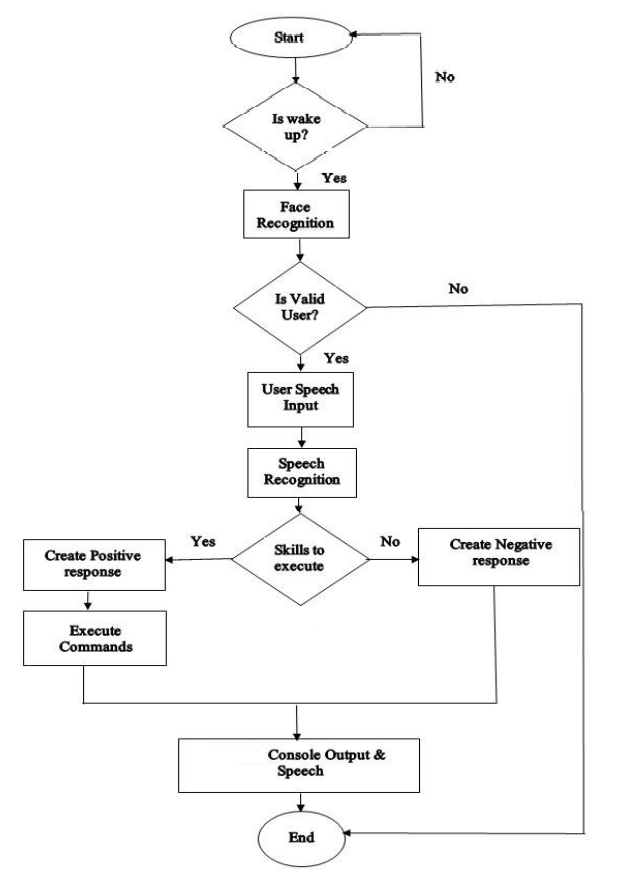
In this situation, there may be several paths to choose from, and it should be able to take into account user preferences, other active tasks, and priorities before recommending one.

****

\*System design and implementation\*

****

\*Data flow Diagram\*

****

\*Flow chart\*

**OBJECTIVES**

* Main objective of building personal assistant software (a virtual assistant) is using semantic data sources available on the web, user-generated content, and providing knowledge from knowledge databases.
* The main purpose of our intelligent virtual assistant is to answer questions that users may have.
* Virtual assistants can tremendously save you time.
* We usually spend hours in online searches and then making the report in our terms of understanding. Our virtual assistant can do that for you. Provide a topic for research and continue with your tasks while it does the research.
* Another difficult task is to open up applications by switching tabs which takes a lot of our precious time. Just tell the assistant what you want to open and see the magic.
* One of the main advantages of voice searches is their rapidity.
* To be a wise voice is reputed to be four times faster than a written search: whereas we can write about 40 words per minute, we are capable of speaking around 150 words during the same period.
* In this respect, the ability of personal assistants to accurately recognize spoken words is a prerequisite for them to be adopted by consumers.

**METHODOLOGY**

An application of a natural language processing-based intelligent voice assistant system that may be used to transmit text messages.  Messages may be sent and received, and the device's other applications can be used on the platform of was designed and implemented utilizing python android.

* **System Architecture**

The System consists of the following stages:

* Collection of data as voice or speech.
* Analysis and conversion of the data into text.
* Storing and processing of the data.
* Generation of speech from the processed data.
* Task Automation on the voiced demand.

The speech and voice samples are gathered as data and saved as input to the following level in the first stage. The data is processed and then turned to text in the second stage. The text is then inspected once it has been converted, and a Python script is used to analyse and identify the command's answer. The output is created in the form of voice from the basic text in the next stage, based on the identified answer.

The aforesaid procedure is then automated utilizing voice commands in the last stage. TTS may be used for everything from basic content to conversational changes.

The following contains the list of features the proposed system will provide:

* + - It continuously listens for its name and will activate as a response after calling.
    - It continues learning the arrangement of inquiries posed to it identified with its setting which it recollects for what's to come. So, when a similar setting is referenced, it begins a discussion with you posing similar inquiries.

* + - It can perform arithmetic calculations upon specific voice commands and can give back the response through voice.
    - It can search the internet based on the user’s commands and reply with either interactive questions or definitive answers through voice.
    - The cloud server to update is a firebase cloud server.
    - It can automatically connect to a cloud server through auto synchronization and can keep itself updated.
    - The architecture based on the Internet of Things can help the user to connect the smart devices.
    - Other highlights, for example, playing music, setting a caution, checking climate states of gadget's area. Setting updates, spell-right, and so forth can be performed by a contribution from the client's voice.
* **Methodology for voice email capability**

 The voice mail capability is enabled and is the key feature to the assistant. Initially if mail command is enabled to assistant by the user the assistant process the requisites of the mail like receiver mail-id, subject and message body of the mail all these requisites are taken as input through voice and it is taken accurately.

Firstly, step is initialized with auto credentials which are fed with assistant automatically so its auto composed. The modules involved in this design are predefined and open-source modules and they are HTTPs requests module, smtplib module which is accompanied with email MIME Handlers which is very useful to holding attachments and also it makes us to take the files from working directory of PC to email through python operating program workspace.

The smtplib module characterizes a SMTP customer session object that can be utilized to send mail to any Internet machine with a SMTP or ESMTP audience daemon. SMTP represents Simple Mail Transfer Protocol. The smtplib modules is valuable for speaking with mail servers to send letters. Sending letters is finished with Python's smtplib utilizing a SMTP server.

Genuine use differs relying upon unpredictability of the email and settings of the email server, the directions here depend on sending email through Gmail. The other two noteworthy fragments of the pack are the "parser" and the "generator". The parser takes the serialized adjustment of an email message (a surge of bytes) and changes over it into a tree of Email Message objects.

The generator takes an Email Message and changes it over into a serialized byte stream. (The parser and generator also handle surges of substance characters, anyway this utilization is incapacitated as it is too easy to even consider evening consider winding up with messages that are not authentic in some way or another.)

The email pack advances a valiant exertion to cover the nuances of the distinctive administering RFCs from the application. Sensibly the application should have the alternative to treat the email message as a composed tree of Unicode substance and twofold associations, without obsessing about how these are addressed when serialized. Before long, nevertheless, it is much of the time essential to think about likely a segment of the standards overseeing MIME messages and their structure, expressly the names and nature of the MIME "content sorts" and how they recognize multipart reports. For the most part this data should simply be required for progressively complex applications, and still, by the day's end it should simply be the huge level structure being alluded to, and not the nuances of how those structures are addressed.

Since MIME substance types are used comprehensively in present day web programming (not just email), this will be a characteristic plan to various programming engineers.

**EXPERIMENTAL SETUP**

The program is made to be lightweight so that it does not put a strain on the system that runs it. This system is being built with widely accessible hardware and software compatibility in mind. The following are the virtual assistant's minimal hardware and software requirements.

**Hardware:**

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

**Software:**

* + - * Windows 7 or later
      * Python 3

Let’s talk a little about the software we are going to use:

* **PYTHON:**

Python is a high-level, interpreted programming language based on OOPs (Object Oriented Programming). It's a powerful, practical language designed for quick application development (RAD).

Python facilitates the writing and execution of programs. When compared to other OOPs languages, Python may achieve the same logic with as little as a fifth of the code.

Python has a long list of advantages for everyone. Python is so versatile that it can't be used for just one thing. Its increasing popularity has enabled it to penetrate some of the most popular and difficult processes, including Artificial Intelligence (AI), Machine Learning (ML), natural language processing, data science, and so on.

Python provides a plethora of libraries to meet this project's needs. Speech recognition for voice recognition.

Python is a relatively efficient programming language. For simple cases, efficiency is typically not an issue.

If your Python code is inefficient, a typical approach to improving it is to figure out what takes the most time and implement that section more effectively in a lower-level language. Writing everything in a low-level language will result in considerably less programming and more efficient code (since you will have more time to optimize).

* **Pyttsx:**

Python Text to Speech is abbreviated as Pyttsx. It's a Python wrapper for text-to-speech synthesis that works across platforms. It's a Python package for Mac OS X, Windows, and Linux that supports standard text-to-speech engines. It works with Python versions 2.x and 3.x. Its key benefit is that it may be used without being connected to the internet.

* **Feasibility Study**

Feasibility study can help you determine whether or not you should proceed with your project. It is essential to evaluate cost and benefit. It is essential to evaluate cost and benefit of the proposed system. Five types of feasibility study are taken into consideration.

1. **Technical feasibility**: It includes finding out technologies for the project, both hardware and software. For virtual assistant, user must have microphone to convey their message and a speaker to listen when system speaks. These are very cheap now a days and everyone generally possess them. Besides, system needs 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 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 to write can read out problems for system and get answers.

3. **Economic feasibility:** Here, we find the total cost and benefit of the proposed system over current system. For this project, the main cost is documentation cost. User also would have to pay for microphone and speakers. Again, they are cheap and available. As far as maintenance is concerned, our virtual assistance won’t cost too much.

4. **Organizational feasibility**: This shows the management and organizational structure of the project. This project is not built by a team. The management tasks are all to be carried out by a single person. 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 cultural environment. Virtual assistant is built in accordance with the general culture.

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. Decision is taken to proceed with the project

**RESULT**

In this research, we created a voice assistant that can perform any task in exchange for requests from users without making any mistakes. More capabilities have been added, such as the ability to listen to the user's voice.

Exclusively, and will not be activated by noise in the environment. This project's modular design makes it simple to modify, understandable and adaptable. We can expand the program's capabilities without affecting its functionality.

All The necessary Python packages have been installed, and the code has been written in Python.

Integrated Development Environment (IDE). The python version used for this project is 3, and the python version used for this project is 3.

Data on various noises were maybe collected from the environment as part of the methodology.

**CONCLUSION AND FUTURE SCOPE**

The desktop assistant's efficiency has been greatly increased in terms of functionality, and it also monitors the user's security using voice recognition technologies.

It's called advanced desktop assistant since it includes a lot of sophisticated capabilities that Microsoft Cortana doesn't have.

With significant features, this assistant is also focused on the user interface, which is necessary because Cortana lacks a user-friendly interface. A new desktop assistant is required to provide next-generation services, allowing us to make desktops as user-friendly as mobile devices.

The Contact’s Application is used to store the client's contact details and also encourages the customer to make a call or send an SMS to other people individual who is using the chemical stored in this app.

* **Future scope:**

Virtual assistants are now available and are quick and responsive, but there is still a long way to go. The present systems' understanding and dependability must be improved. It has vastly improved.

Today's aides are still insufficient. In critical situations, it's dependable. These helpers' future will be bright. Artificial Intelligence (AI) should be implemented into virtual assistants. Machine Learning and Neural Networks are examples of intelligence. IoT and networks, for example.

With the addition of these, we shall be able to reach new heights thanks to technological advancements. What really is the situation?

What virtual assistants can accomplish is far beyond what we now have, achieved up until now Jarvis is a virtual reality that most of us have seen.

**REFERENCES**

* **Websites referred:**
* www.stackoverflow.com
* www.pythonprogramming.net
* www.codecademy.com
* www.tutorialspoint.com
* **Books Referred:**
  + Python Programming -Kiran Gurbani
  + Learning Python – Mark Lutz
  + Hands-on Machine Learning with Scikit-Learn (O’Reilly)
* **YouTube Channels Referred:**
  + Code with Harry
  + Edureka!
  + Programming with Mosh
* **Documents Referred:**
  + C.K. Gomathy. (2010),"Cloud Computing: Business Management for Effective Service Oriented Architecture" International Journal of Power Control Signal and Computation (IJPCSC), Volume 1, Issue IV, Oct - Dec 2010, P. No:22-27, ISSN: 0976-268X
  + “VIRTUAL PERSONAL ASSISTANT (VPA) FOR MOBILE USERS”
  + D. SOMESHWAR, DHARMIK BHANUSHALI, SWATI NADKARNI, “IMPLEMENTATION OF VIRTUAL ASSISTANT WITH SIGN LANGUAGE USING DEEP LEARNING AND TENSORFLOW”
  + C K Gomathy and V Geetha. Article: A Real Time Analysis of Service Based Using Mobile Phone Controlled Vehicle Using DTMF For Accident Prevention. International Journal of Computer Applications 138(2):11-13, March 2016. Published By Foundation of Computer Science (FCS), NY, USA, ISSN No: 0975-8887