

“SKYE : Voice Based AI Desktop Assistant”

Hardik Muley¹, Jaydeep Ghosh², Ankit Lal Sinha³, Prof. Padmavati Sarode⁴

^{1,2,3} Student, Department of Computer Science & Engineering, GHRCEM, Pune, Maharashtra, India

⁴ Professor, Department of Computer Science & Engineering, GHRCEM, Pune, Maharashtra, India

Abstract - Nowadays, the field of technology is rapidly advancing with each passing day. In the past, computers were only capable of performing limited tasks, but now, with the emergence of new technologies such as machine learning, artificial intelligence, deep learning, and a few others have advanced computer systems to the point where we can perform any sort of task with them. One of the popular technology, artificial intelligence (AI), is to achieve a natural dialogue between humans and machines. This paper demonstrates how our suggested voice assistant, which offers aid to individuals which are Visually impaired and with disabilities. A voice assistant refers to a software application that utilizes voice technology to comprehend spoken language and generate artificial vocal responses. The goal is to create a Python-based desktop voice assistant that can assist users with a wide range of tasks without requiring them to interact with a keyboard. The purpose of this project is to explore the intelligent behavior of voice assistants and their potential applications for both daily use and education.

Key Words: Python, speech recognition, voice assistant, TTS (Text to Speech), STT(Speech to Text), Desktop Assistant, Artificial Intelligence.

1. INTRODUCTION

In today's fast-paced world, we often find ourselves wanting to complete tasks more efficiently by using voice commands instead of relying solely on a keyboard and mouse. This is particularly useful for multitasking, and companies such as Google, Amazon, and Apple are working to make this technology more widely available. For instance, we can easily set reminders or alarms using voice commands. With this in mind, we have developed a platform that can be installed in any location and used to assist people with various tasks through voice-based communication.

A virtual assistant is an artificial intelligence technology designed to help users with basic tasks by responding to natural language commands. The system works by converting audio signals into digital data, which is then analyzed by the software. This personalized speech recognition project is built on Python and is capable of recognizing user commands, interacting with users, and completing tasks accordingly. Examples of tasks it can perform include greeting users based on the time of day or event, playing music, providing weather updates, analyzing weather conditions and advising users on whether it's safe to go out, opening applications and folders, creating new folders, changing directories, sending emails, and more. Using a virtual assistant to perform these tasks can save users a significant amount of time and effort.

Focusing on what is most important, whether it is personal or professional, is crucial. However, individuals often spend too much time on mundane tasks that can be automated with personal assistants. When people are not familiar with a work environment, they may have trouble locating necessary applications such as browsers, IDEs, or other software. This can result in hours wasted searching for applications and ultimately wasted time. To mitigate this issue, a voice-enabled personal assistant can be employed to automate this process. By simply issuing a voice command, the assistant will take care of the rest. Artificial intelligence-based voice assistants have numerous applications, including IT Helpdesk, home automation, HR-related tasks, and voice-based search. Voice-based search is expected to be the future of next-generation technology, with users relying heavily on voice assistants for all their needs. These assistants will also be particularly useful for visually impaired individuals. Worldwide, 15% of the population has some form of disability, with 2-4% experiencing significant difficulties with their daily functioning. Using a website can be challenging for people with disabilities, so we aimed to develop a unique way for people with different needs to access the internet. To create our project, we used Visual Studio Code, as well as modules and libraries such as pyttsx3, Speech Recognition, Date time, Wikipedia, Smtplib, and more.

2. LITERATURE REVIEW

In modern times, we teach our machines to emulate human thinking and perform tasks independently, resulting in machines replacing human labor. As a result, the concept of voice assistants has emerged, which can accomplish various tasks for humans based on voice commands. The virtual assistant has the ability to comprehend and sort out particular directives provided by the user, providing pertinent information in response.

Voice control is becoming a new way for humans to interact with machines, where analog signals are converted into digital waves through speech signals. Smartphone usage has skyrocketed in recent years, resulting in extensive use of voice assistants such as Apple's Siri, Google's Assistant, and Microsoft's Cortana, and Amazon's Alexa. These voice assistants are created utilising technologies like as voice recognition, speech synthesis, and Natural Language Processing (NLP) to provide consumers with a variety of applications that make their life easier and more pleasant.

Deepak Shende and Ria Umabiya stated that AIVA, which was introduced in 2018, is an intelligent assistant, along with Microsoft's Cortana and Google Assistant from Google. Their objective is to create a personal assistant that is controlled by voice commands and can perform a variety of tasks, including conducting Internet searches and posting comments on social media platforms such as Facebook and Twitter. The assistant also has the capability to provide weather updates for the user's region. Simple voice instructions can readily achieve these things. [5]

C. Vimala and V. Radha stated that among humans, speech is the most prevalent mode of communication. Automated speech recognition has become popular because humans tend to prefer machines that can communicate using speech, which is considered the most sophisticated method. DTW and HMM are the most commonly used speech recognition techniques. MFCC, which provides distinct dimensions of the sound signal, is used for speech feature extraction. Previous research has shown that MFCC is more accurate and realistic than other techniques used for mining voice characteristics. The study was conducted in MATLAB, and the results indicate that the machine is capable of detecting words with a high level of accuracy. [4]

Dr. Kshama V. Kulhalli conducted a survey to compare the performance of the top voice assistants, namely Google assistant, Apple's Siri, and Microsoft's Cortana. The survey results indicated that Google assistant provided the most accurate responses compared to the other assistants. Google assistant was able to easily recognize variations in voice. [7]

Speech technology is a popular and versatile technology that can be used for various applications. It enables robots to interact with humans in a structured and appropriate way, providing useful services. The study covers the fundamentals of the speech recognition process and its various models and applications, as well as a description of the ongoing research into the various techniques used in speech recognition systems. Speech recognition systems continue to advance and have limitless applications. [6]

Tulshan suggested that frequent typing can lead to finger injuries. To prevent such issues, we require a system that can perform tasks based on voice commands. The system will recognize the user's voice, process the recognized words, and display them on the screen if they are relevant and meaningful. Following this, specific keywords will be recognized to compile and execute the program. [3].

3. TECHNOLOGIES USED

a. Python:

- Python is a well-liked programming language at a high level that is recognized for its straightforwardness, comprehensibility, and user-friendliness.
- Python 3.10.0 is being utilised in the development of the voice assistant project.

b. Visual Studio Code:

- Microsoft has created a well-known open-source code editor called Visual Studio Code, often referred to as VS Code.
- VS Code supports a wide range of programming languages and offers features such as syntax highlighting, code completion, debugging, Git integration, and extensions that can be installed to add more functionality.

4. METHODOLOGY

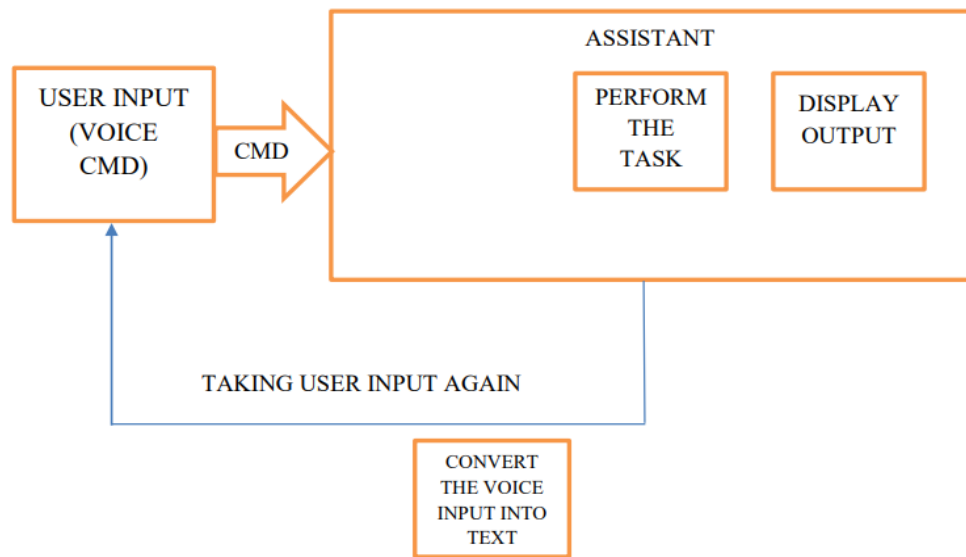


Fig 1: basic workflow of model

There are three modules in this Assistant. The first step is for the assistant to take user voice input. Second, analyse the user input and translate it to the appropriate intent and function. The third is the assistant providing the user with the outcome all along through speech. The assistant will first begin receiving human input. When the assistant receives the input, it will transform the analogue voice input to digital text. If the assistant is unable to turn the voice into text, it will prompt the user for input once more. After converted, it will begin processing the input and mapping it to a certain function. The output will then be provided to the user by voice command. Users request or inquiry gets chops into segregate commands which makes it easy to recognize by our desktop voice assistant.

- In comparison to other inquiries, ours is searched inside the command list
- The voice assistant receives these orders through the command list.
- Once the voice assistant accepts or receives a command, it will immediately determine the appropriate action to take.
- If the user's inquiry is not understandable, then the voice assistant will ask for clarification before proceeding.
- Particularly, the voice assistant detects what we want to get.
- When the voice assistant recognises the command and determines that it can proceed, it will provide the person or user with the necessary information.

For example, When a person says, "SKYE, open Whatsapp or wikipedia," the voice assistant will listen to the command and take the appropriate action, such as opening the relevant website. After the user has finished speaking, the voice assistant will pause for a few seconds to ensure it has captured the entire request, and then it will search its database for the inquiry to provide the relevant result.

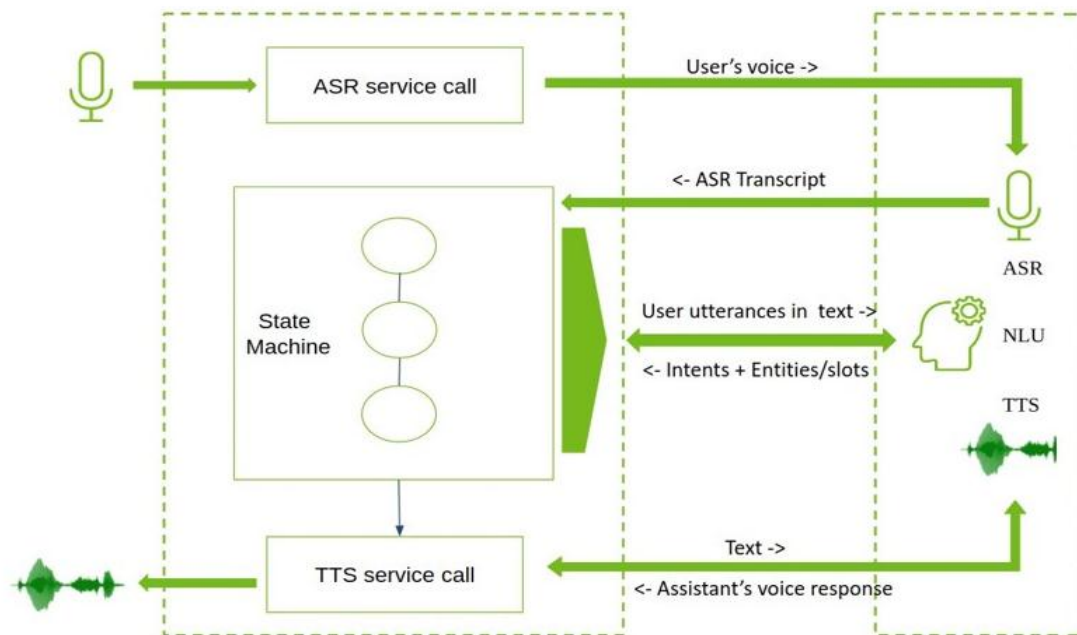


Fig 2: detailed workflow of model

The following are several modules employed to create a voice assistant capable of performing various common functions.

```

1  import pyttsx3 #pip install pyttsx3
2  import speech_recognition as sr #pip install speechRecognition
3  import datetime
4  import os
5  import wikipedia
6  import random
7  from requests import get
8  import webbrowser
9  import pywhatkit as kit
10 import smtplib #pip install secure-smtplib
11 import sys
12 import time
13 import pyjokes
14 import PyPDF2
15 import pyautogui
16 import requests

```

Fig.3. different packages used

4.1 Packages used :

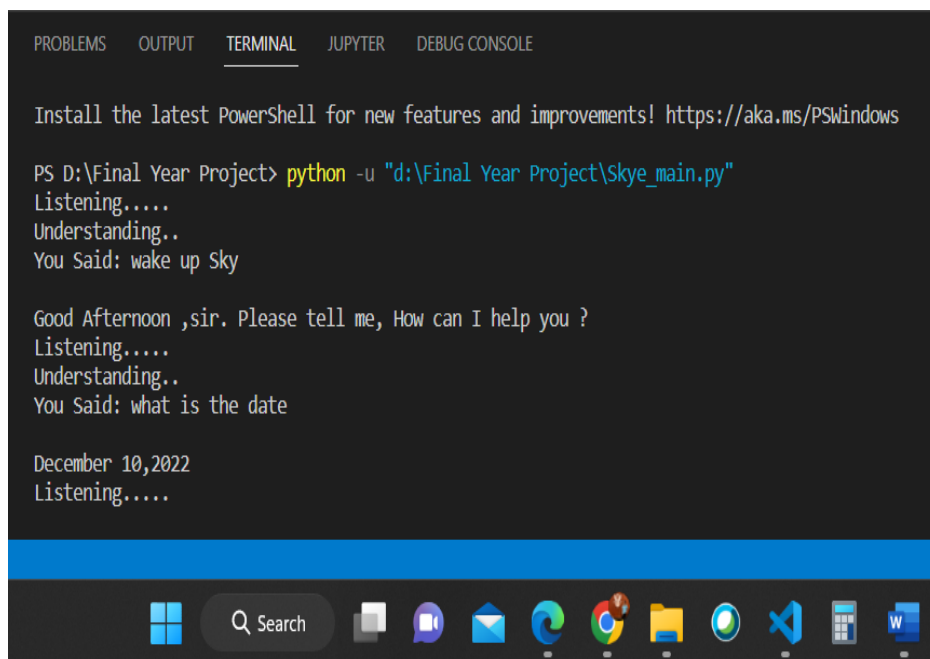
- a. **Speech:** RecognitionThe Speech Recognition library is employed to capture spoken words from a microphone and process them to determine their meaning and convert them into text format. This library enables machine systems to comprehend and interpret human language.

- b. **Pytttsx3:** The Pytttsx3 library, which stands for Python text to speech, is utilized to enable our voice assistant to communicate with us audibly. It supports various text to speech engines that can convert text into speech, allowing the voice assistant to speak to its user. We can select the voice to be male or female based on our preferences.
- c. **Wikipedia:** To retrieve information from Wikipedia on any topic, perform a search, or seek solutions to a query, we must employ the Wikipedia library. This Python library requires an internet connection to obtain results, and it can present the findings to the user in both text and voice format.
- d. **Datetime:** This module is crucial for facilitating date and time-related functionalities. It is useful when a user needs to access the current date and time or when they want to schedule a task at a specific time.
- e. **PyWhatkit:** PyWhatKit is a Python library that offers several functionalities, including sending messages and images via WhatsApp, playing YouTube videos, converting images to ASCII art, sending emails, and more.
- f. **OS (Operating System):** The OS module in Python is utilized for interacting with the operating system. Specifically, we can use the 'Start file()' function to launch any installed application on our system.
- g. **webbrowser:** The webbrowser module in Python offers a user-friendly way to open and show web pages in a web browser window. It provides the option to open a web page in either a new browser window or the current one, depending on your preferences.

5. RESULTS

a. Asking for today's date

As shown in Fig.4. We have asked skye for the current date and he told us the current date.



```

PROBLEMS  OUTPUT  TERMINAL  JUPYTER  DEBUG CONSOLE

Install the latest PowerShell for new features and improvements! https://aka.ms/PSWindows

PS D:\Final Year Project> python -u "d:\Final Year Project\Skye_main.py"
Listening....
Understanding..
You Said: wake up Sky

Good Afternoon ,sir. Please tell me, How can I help you ?
Listening....
Understanding..
You Said: what is the date

December 10, 2022
Listening....
  
```

Fig.4. Output of screen for displaying date

b. Opening Google

As shown in Fig.5. We have asked skye to Open Google. It receives the request and performs action on it

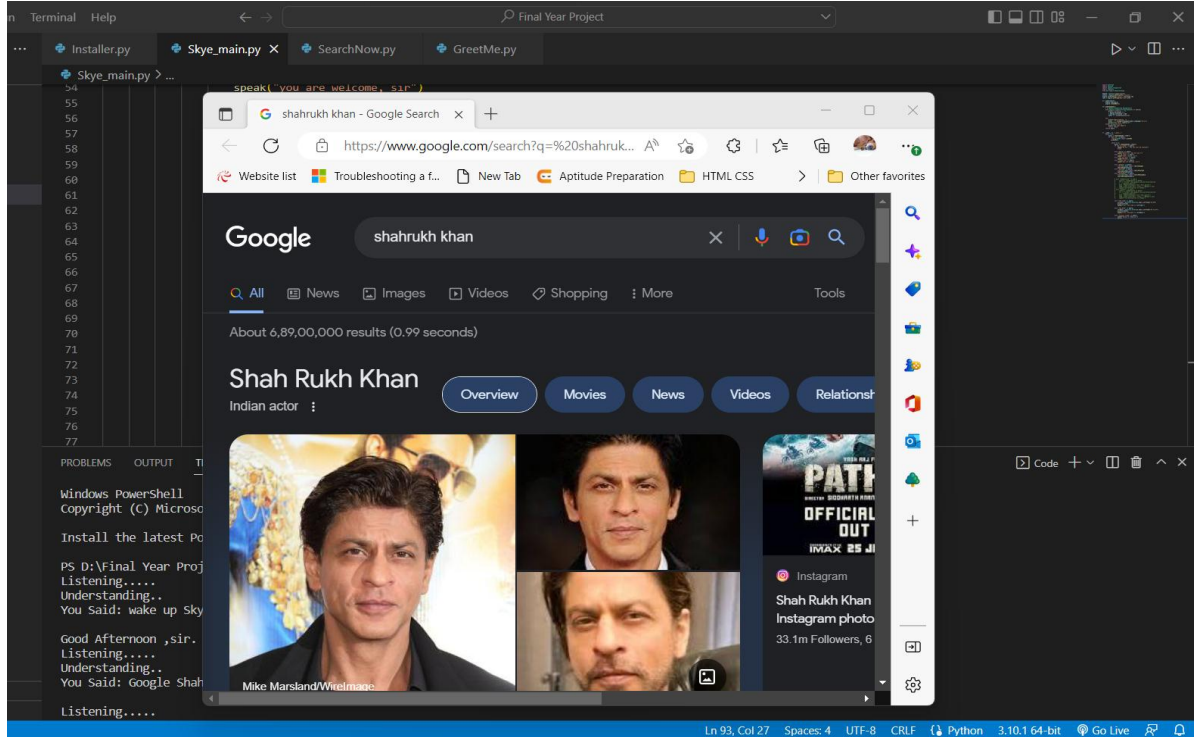


Fig.5. Output of screen for Google search

c. Opening youtube and searching tom and jerry

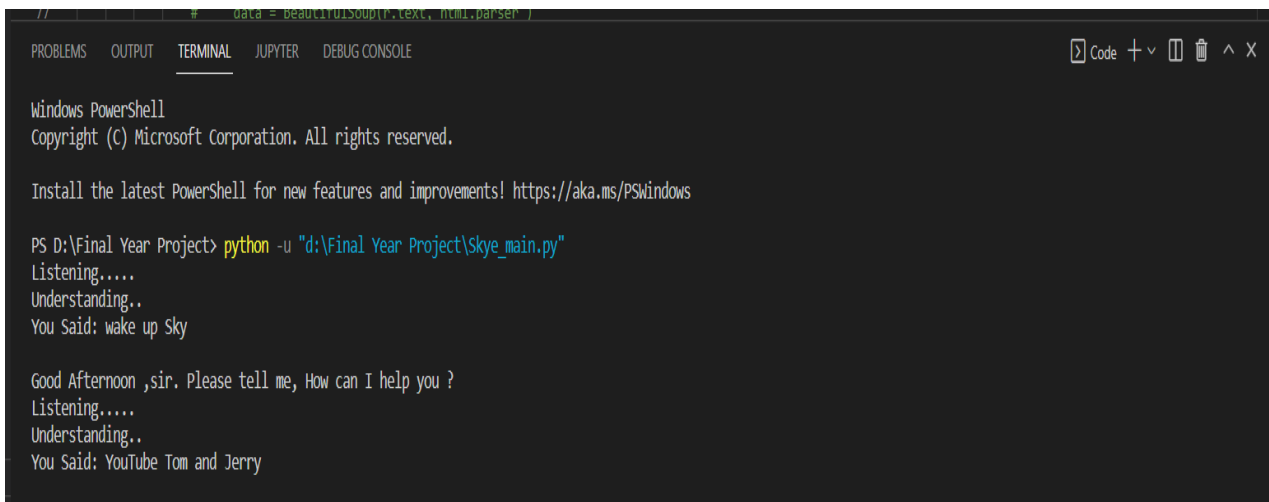


Fig.6. Output of screen for opening youtube

As shown in Fig.7. We have asked skye to Open Youtube and search tom and jerry video on youtube skye receives the request and performs the given task.

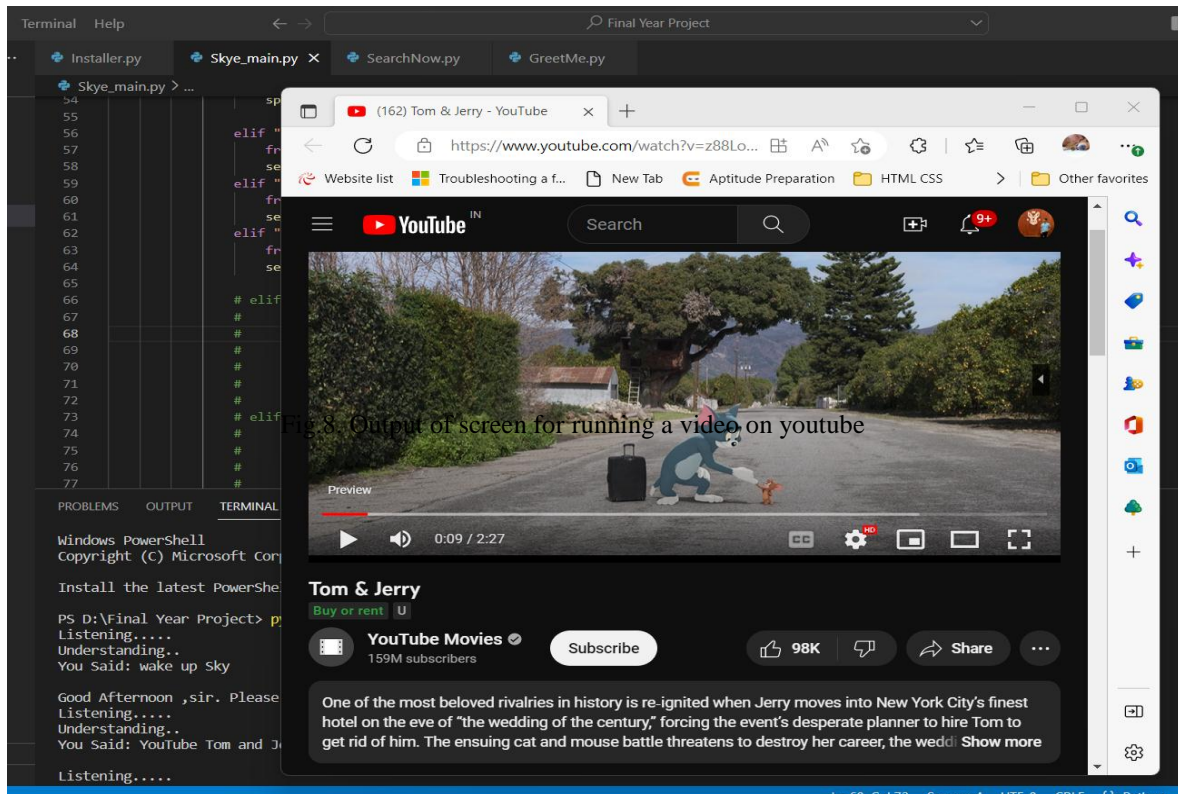


Fig.7. Output of screen for running a video on youtube

6. FUTURE SCOPE

At present, the program is restricted to the English language; however, there are intentions to broaden its accessibility to other languages shortly. The goal is to create a reliable software that requires minimal typing and can be operated entirely through voice commands, providing a seamless user experience. To ensure widespread adoption, it's important to minimize the software's reliance on the local environment and operating system.

Our virtual assistant will soon have a read-aloud function that will allow individuals with disabilities to listen to and access desired information from various web resources. Currently, this feature is only available on PC, but in the future, it will be available on all devices. Additionally, the feature will be designed with ease of use and user-friendliness in mind for individuals with disabilities, ensuring that they do not need constant supervision to use it effectively.

7. CONCLUSIONS

AI-powered voice assistants for desktops have transformed the way we interact with our computers. By utilizing advanced algorithms and technologies, these assistants can understand and interpret user speech, allowing for more efficient and effortless task performance. With a broad range of functions such as messaging, calling, and playing music, these voice assistants have become an indispensable tool for many users. As technology progresses, we can anticipate these voice assistants to become even more sophisticated, further simplifying our lives. This project will benefit individuals of all ages and those with disabilities or unique circumstances. The personal voice assistant will be user-friendly and minimize the need for manual human efforts to accomplish various tasks. The current voice assistant system operates exclusively on desktops. However, the modular nature of the system allows for additional features to be added without disrupting the current system functionalities.

8. REFERENCES

- [1] "ASKI The Virtual Desktop AI-Based Voice Assistant", ISSN 2581-9429 Volume 02, Number 1 (2022).
- [2] "Desktop voice guide using python and Artificial Intelligence," ISSN 2582-5208 Volume 04, Number 5 (2022).
- [3] Tulshan, Amrita & Dhage, Sudhir. (2019). "Survey on Virtual Assistant: Google Assistant, Siri, Cortana, Alexa", 4th International Symposium SIRS 2018, Bangalore, India, September 19–22, 2018, Revised Selected Papers. 10.1007/978-981-13-5758-9_17.
- [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] Deepak Shende. Ria Umabiya, Monika Raghorte, Aishwarya Bhisikar. Anup Bhangre. "AI Based Voice Assistant Using Python", International Journal of Emerging Technologies and Innovative Research (www.jetir.org), ISSN 2349-5162, Vol.6, Issue 2, page no.506-509, February-2019.
- [6] Srivastava S., Prakash S. (2020) Security Enhancement of IoT Based Smart Home Using Hybrid Technique. In: Bhattacharjee A., Borgohain S., Soni B., Verma G., Gao XZ. (eds) Machine Learning, Image Processing, Network Security and Data Sciences. MIND 2020. Communications in Computer and Information Science, vol 1241. Springer, Singapore. https://doi.org/10.1007/978-981-15-6318-8_44.
- [7] Dr. Kshama V. Kulhalli, Dr.Kotrappa Sirbi, Mr. Abhijit J. Patankar, "Personal Assistant with Voice Recognition Intelligence", International Journal of Engineering Research and Technology. ISSN 0974-3154 Volume 10, Number 1 (2017).
- [8] "Comparative Analysis of Smart Voice Assistants" IEEE International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS) 2021.