Personal Desktop Voice Assistant Research Paper

Sakshi R Jain[1], Prof Feon Jason[2]

Master Student, School of CS & IT, Jain University, Bengaluru, India[1]

Professor, School of CS & IT, Jain University, Bengaluru, India[2]

Abstract: A virtual assistant is a software agent that can perform tasks or services for an individual. Sometimes the term "chatbot" is used to refer to virtual assistants generally or specifically those accessed by online chat. Virtual Assistant (VA) is a term that applies to computer-simulated environments that can simulate physical presence in places in the real world, as well as in imaginary worlds. This report discusses ways in which new technology could be harnessed to create an intelligent Virtual Personal Assistant (VPA) with a focus on user-based information. This project is a technical brief on Virtual Assistant technology and its opportunities and challenges in different areas. The project focuses on virtual assistant types and structural elements of a virtual assistant system. In this project, we tried to study virtual Environment and virtual Assistant Interfaces, and the paper presents applications of virtual assistant that helps in providing opportunities for humanity in various domains. This project also describes the challenges of applying virtual Assistant technology. In today's advanced hi-tech world, the need of independent living is recognized in case of visually impaired people who are facing main problem of social restrictiveness. They suffer in strange surroundings without any manual aid. Visual information is the basis for most tasks, so visually impaired people are at disadvantage because necessary information about the surrounding environment is not available. With the recent advances in inclusive technology, it is possible to extend the support given to people with visual impairment.

I.   INTRODUCTION

Computers have become a very important devices and as well as less expensive over time. The idea behind the Personal Virtual Assistant is creation of an inexpensive, reliable and easy to use assistant. Virtual Assistant (VA) is a term that applies to computer-simulated environments that can simulate physical presence in places in the real world, as well as in imaginary worlds. Virtual Assistant is a real-time and interactive technology. It means that the computer can detect user inputs and modify the virtual world instantaneously. Interactivity and its captivating power contribute to the feeling of being part of the action on the environment that the user experience. All human sensorial channels can be used to have a high-level interaction. Most current virtual assistant environments are primarily visual experiences, displayed either on a computer screen, but some simulations include additional sensory information, such as sound through speakers or headphones. Virtual Assistant technology has been a promising technology applicable in various domains of application such as training simulators, medical and health care, rehabilitation, education, engineering, scientific visualization, and the entertainment industry. The application works similar like Siri/ Google Assistant. But the application deals with the computer itself mainly. A voice assistant is a digital assistant that uses voice recognition, natural language processing and speech synthesis to provide aid to users through devices and voice recognition applications. This project is based on basic concepts of Artificial Intelligence such speech recognition.

II.   APPLICATION OF PERSONAL DESKTOP VOICE ASSISTANT

Voice assistants can do the following basic functions:

o Web search
o Play music or videos
o Set reminders and alarms
o Launch any programme or application
o Receive weather updates o Send emails, WhatsApp, etc.

These are only a few instances of the jobs that voice assistants can complete; we can conduct a great deal more depending on our needs. Voice assistants' capabilities and advancements are always growing day by day to give users improved performance. Our desktop-based voice assistant is built using Python modules and libraries, allowing it to function quickly and efficiently on the desktop.

The fundamental premise of our project is that the user asks the voice assistant to complete their task using the device's microphone, and the command is subsequently translated into text. The text request is processed after that, and a text response is provided along with any voice assistant work. In addition to fundamental daily functions, we are attempting to integrate the idea of face identification for security purposes in our voice assistant to give it more versatility and personality. Our application uses the fewest system resources, which decreases the need for expensive systems and lessens the risk to your system because it doesn't communicate with servers directly.

There are many reasons that make this vocal voice command application necessary in practical settings. These are a few of them:-

To make it possible for a very engaging user experience: Like no other interface, voice help keeps users' attention. Users can ask for anything they want by speaking naturally to the programmes.

To eliminate user annoyance with the application: With the current machine system, we must touch, type, and use a mouse to complete our task, which occasionally causes user frustration. Users can ask their desired task directly utilising a voice assistant.

Voice assistants are truly able to answer for each user: based on their location, language, and preferences. This allows you to customise your app experience for each user.

III.   TECHNOLOGY USED IN PERSONAL DESKTOP VOICE ASSISTANT

The desktop voice assistant is a software system that enables users to perform various tasks using voice commands. The system should be able to recognize voice commands accurately, respond promptly, and perform the requested tasks efficiently.

Python: Python is a popular programming language used for developing personal desktop voice assistants. It offers several libraries and frameworks for implementing speech recognition, natural language processing, and machine learning.

Speech recognition APIs: Popular speech recognition APIs include Google Cloud Speech-to-Text API, Amazon Transcribe, and Microsoft Azure Speech Services. These APIs provide speech-to-text conversion capabilities and can be integrated into voice assistant applications.

Natural Language Processing (NLP) libraries: There are several NLP libraries available for Python such as Natural Language Toolkit (NLTK), spaCy, and Stanford CoreNLP. These libraries help with tasks such as sentiment analysis, named entity recognition, and part-of-speech tagging.

Text to Speech (TTS) engines: Popular TTS engines include Google Text-to-Speech, Amazon Polly, and Microsoft Speech Services. These engines can be used to generate human-like speech output from text input.

Machine learning frameworks: Popular machine learning frameworks include TensorFlow, PyTorch, and Scikit-learn. These frameworks can be used to train machine learning models for speech recognition and NLP tasks.

Graphical User Interface (GUI) libraries: GUI libraries such as PyQt and Tkinter can be used to create a visual interface for the voice assistant application. The GUI can be used to display information such as weather updates, news articles, and reminders.

Web APIs: Web APIs such as OpenWeatherMap, NewsAPI, and Spotify Web API can be used to integrate the voice assistant with external services. These APIs allow the voice assistant to access weather forecasts, news articles, and music streaming

Functional Requirements:

The desktop voice assistant should have the following functionalities:

- **Wake word detection:** The system should be able to detect a wake word such as "Hey, assistant" to activate the assistant.
- **Voice recognition:** The system should be able to accurately recognize and interpret voice commands from the user.
- **Natural language processing:** The system should be able to understand the user's intent and respond accordingly.

- **Task execution:** The system should be able to execute tasks such as setting reminders, playing music, sending emails, and searching the web.
- **Multi-language support**: The system should be able to recognize and interpret voice commands in multiple languages.

Non-functional Requirements

The desktop voice assistant should meet the following quality attributes:

- **Accuracy:** The system should have a high accuracy rate in recognizing voice commands.
- **Response time:** The system should respond promptly to user commands.
- **Security:** The system should be secure and protect user privacy.
- **Usability:** The system should be easy to use and have a user-friendly interface.
- Constraints. The desktop voice assistant should work on Windows, Mac, and Linux operating systems. The system should also be compatible with various microphones and audio input devices.
- **Assumptions and Dependencies:**
- The desktop voice assistant assumes that the user has a reliable internet connection for certain tasks such as web searches and email sending. The system also depends on third-party APIs for certain functionalities such as weather updates and music streaming.

 Requirements:

1.   **Software Requirements:**
- Windows OS.

2.   **Hardware Requirements:**
- Minimum Requirement – 2 Gb RAM, Microphones.
- Recommended – 4 Gb RAM, Microphones.

Other Requirements:
- Internet Connection.

IV.   RESEARCH METHODOLOGY

We are going to use python language and google text to speech API for this project, speech recognition module can be used to recognize the voice of used, and based on its query will be fired. Many different modules i.e., web browser, YouTube, Wikipedia, etc. are used to interact with the internet. the os module is used to interact with operating system related queries.

For Learning purposes, users can search any information related to a certain topic on Wikipedia, Google or in text documents. We are using some concepts related to AI and NLP for the processing of text into voice. Our project's main goal is to develop a virtual voice assistant for blind people so they may use it to communicate with emerging technologies, manage their devices, and learn from them.

V.   PROBLEM STATEMENT

Create a virtual voice assistant, so that the system can be used by them to interact with new technologies, control their devices and use this technology for educational purposes. This is an innovative System for visually impaired people and acts as a voice assistant for them. This system is used to help the visually impaired to have access to the most important features of the desktop enhancing the quality of the system making use of different custom layouts and using speech to text. All actions performed by the system will be based on voice of the user.

The System helps the user on the basis of voice note i.e. the system works on command given by the user. The System speaks out if anything needs to be replied back to the user as user can't see the action taking place in the desktop.

• The blind candidate will also feel self-sufficient.
• As the system is a machine, it won't make mistakes during the execution.
• It will control your device just by voice commands.
• The assistant will understand the context and give the user response accordingly.
•There are many senior citizens who can't use desktop but this will also be useful for them.

These kinds of assistive technologies will help visually impaired users to learn, compete and communicate with their sighted peers.

VI.   PROPOSED METHODOLOGY

- A personal desktop voice assistant is a software application that is designed to understand and respond to voice commands provided by the user. The following is a high-level system design description for a personal desktop voice assistant.

- User Interface: The user interface of the personal desktop voice assistant should be intuitive and easy to use. The user should be able to interact with the voice assistant through natural language commands.

- Speech Recognition: The system should have a robust speech recognition module that can accurately convert the user's voice commands into text. The speech recognition module should also be able to distinguish between different users and adapt to their speech patterns.

- Natural Language Processing (NLP): The system should have an NLP module that can interpret the user's text commands and extract the relevant information. The NLP module should also be able to identify the user's intent and provide appropriate responses.

- Knowledge Base: The system should have a knowledge base that contains information on a wide range of topics. The knowledge base should be regularly updated to ensure that the voice assistant can provide accurate and up-to-date information.

- Machine Learning: The system should use machine learning algorithms to continuously improve its performance. The machine learning algorithms can be used to improve speech recognition accuracy, NLP performance, and user interaction.

- APIs: The system should be able to integrate with other applications and services through APIs. This will enable the voice assistant to provide more comprehensive responses to user requests.

- Privacy and Security: The system should be designed to protect user privacy and ensure that user data is secure. The system should only collect data that is necessary to provide the voice assistant's services, and user data should be encrypted and stored securely.

- User Personalization: The system should be able to personalize the user experience based on the user's preferences and previous interactions. The system should also be able to learn from user feedback and adapt to their preferences over time.

- Action Fulfillment: Once the intent of the user's request is identified, the system will execute the necessary action. For example, if the user requested to play a song, the system will find the song and play it through the desktop speakers.

- Response Generation: Finally, the system will generate a response to confirm that the requested action has been completed. The response could be as simple as a confirmation message, or it could be more detailed, providing additional information related to the user's request.

- Overall, a personal desktop voice assistant should be designed to provide a seamless and natural interaction between the user and the system. The system should be reliable, accurate, and secure, and should continuously learn and improve to provide better services to the user.

Workflow Diagram:
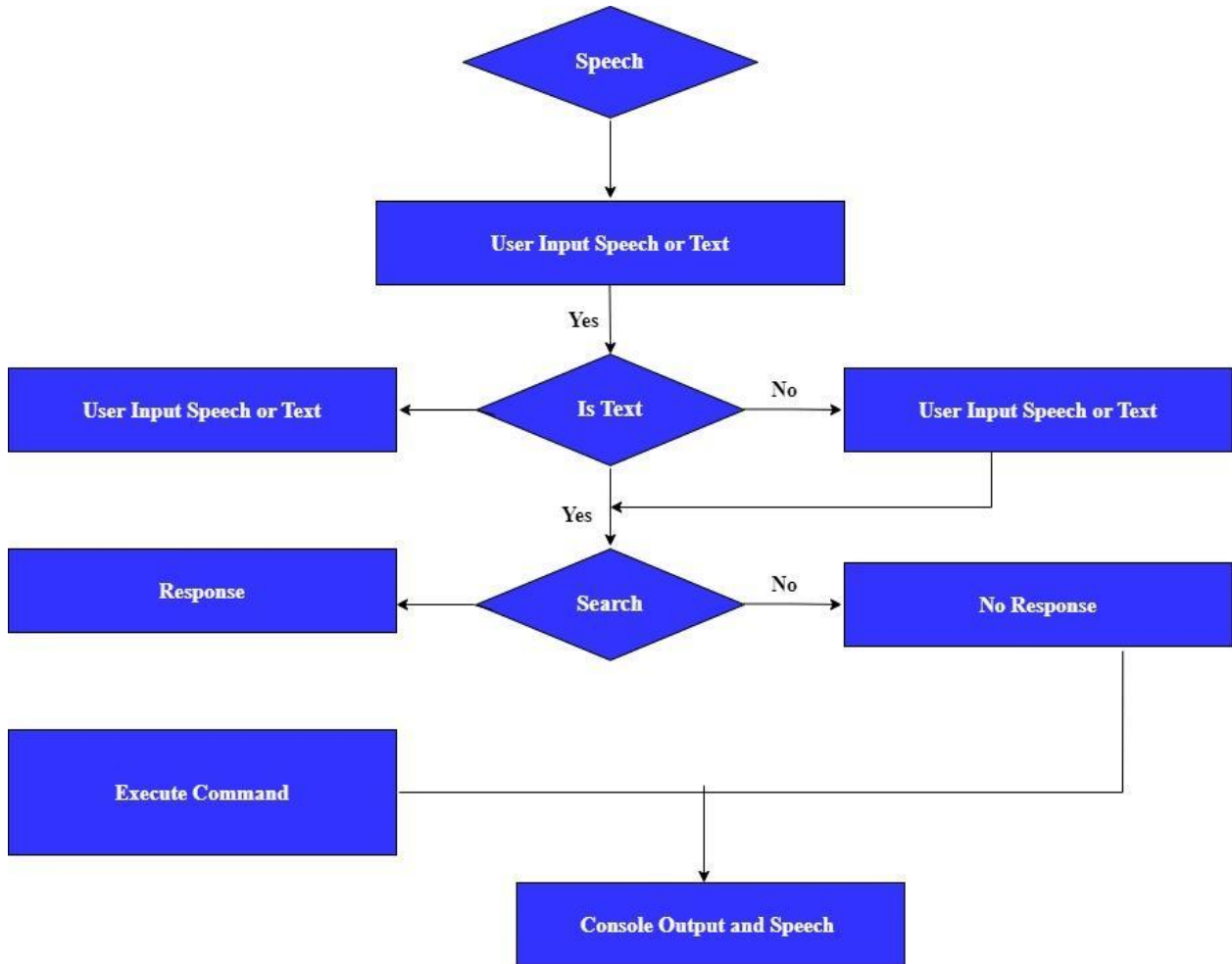
Wake word detection: The voice assistant is always listening for a specific wake word or phrase (e.g. "Hey Siri" or "Okay Google"). When the user says this wake word, the assistant wakes up and starts listening for the user's request.

Speech recognition: The user speaks their request, which is then captured by the assistant's microphone and converted to text using speech recognition technology.

Natural Language Understanding (NLU): The assistant analyzes the user's request to understand the intent behind it. This involves identifying the keywords and context of the request to determine the user's desired action.

Intent mapping: The assistant maps the user's request to a specific action or set of actions. For example, if the user asks "What's the weather like today?", the assistant would map this intent to a weather app and retrieve the relevant information.

Action execution: The assistant executes the mapped action(s) and retrieves the requested information, which is then presented to the user through text or speech.

Response generation: The assistant generates a response to the user's request, which is either spoken aloud or displayed on the screen. This response may include the requested information, confirmation of a completed action, or an error message if the assistant was unable to fulfill the request.

End of session: Once the assistant has provided a response, it goes back to "sleep" and waits for the next wake word to begin a new session.

Use-Case Diagram:
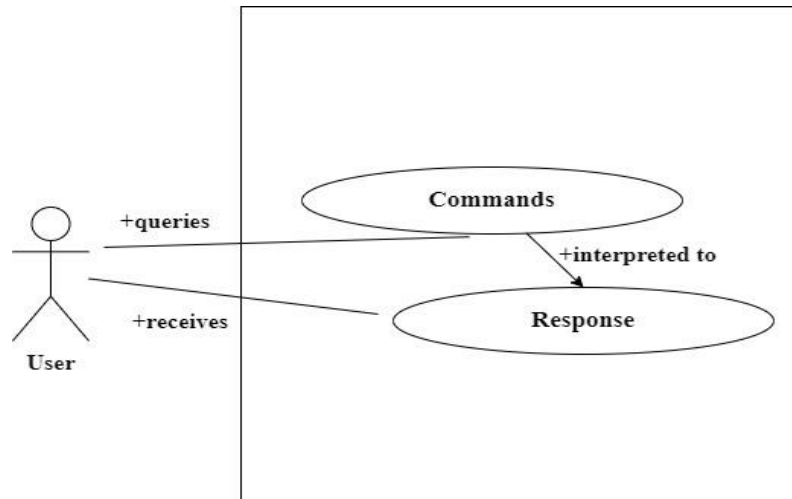
The use case diagram shows the main interactions between the user and the personal voice assistant. The user initiates the interactions by speaking to the assistant, which responds with the requested information or action. The as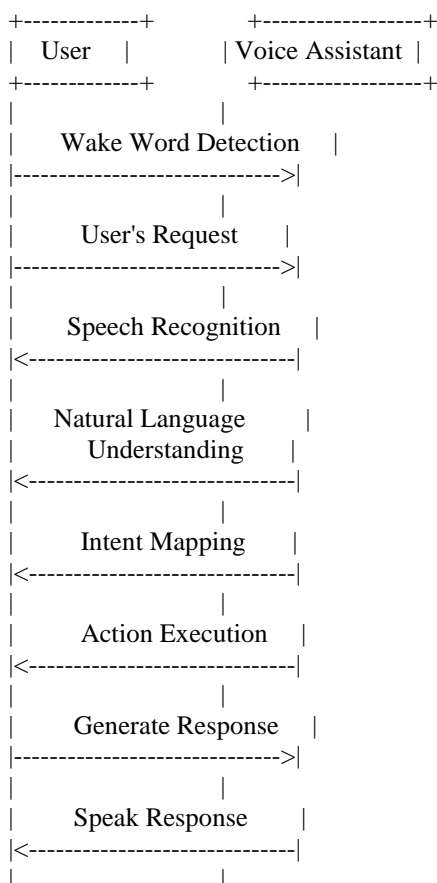sistant is responsible for recognizing the user's speech, interpreting their intent, executing the relevant action, and generating an appropriate response.

Some of the common use cases for a personal voice desktop assistant include asking for information, setting reminders or appointments, sending emails, controlling home automation devices, playing music, performing web searches, and checking the weather or news updates.

Sequence Diagram:

```
+-------------+          +-----------------+
|   User     |          | Voice Assistant |
+-------------+          +-----------------+
|             |                 |
|    Wake Word Detection    |
|----------------------------->|
|             |                 |
|      User's Request     |
|----------------------------->|
|             |                 |
|    Speech Recognition    |
|<----------------------------|
|             |                 |
|    Natural Language      |
|      Understanding       |
|<----------------------------|
|             |                 |
|     Intent Mapping      |
|<----------------------------|
|             |                 |
|     Action Execution    |
|<----------------------------|
|             |                 |
|    Generate Response    |
|----------------------------->|
|             |                 |
|     Speak Response      |
|<----------------------------|
|             |                 |
```
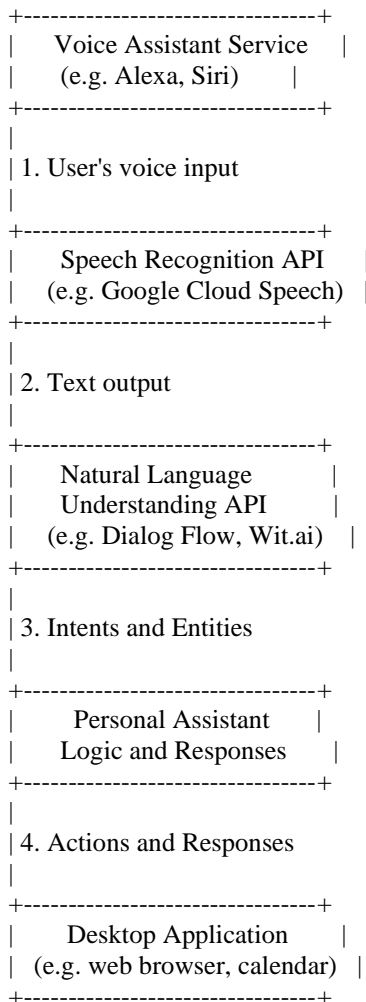
The sequence diagram shows the interaction between the user and the voice assistant. When the user speaks the wake word, the assistant wakes up and starts listening for the user's request. The user then speaks their request, which is captured by the assistant's microphone and converted to text using speech recognition technology. The assistant analyzes the user's request to understand the intent behind it and maps it to a specific action or set of actions. The assistant then executes the mapped action(s) and generates a response to the user's request. Finally, the assistant speaks the response to the user.

This sequence diagram illustrates the flow of communication between the user and the voice assistant and highlights the key steps involved in processing the user's request.

Architecture Diagram:

```
+-------------------------------+
|    Voice Assistant Service    |
|     (e.g. Alexa, Siri)        |
+-------------------------------+
|
| 1. User's voice input
|
+-------------------------------+
|    Speech Recognition API     |
|   (e.g. Google Cloud Speech)  |
+-------------------------------+
|
| 2. Text output
|
+-------------------------------+
|    Natural Language           |
|    Understanding API          |
|   (e.g. Dialog Flow, Wit.ai)  |
+-------------------------------+
|
| 3. Intents and Entities
|
+-------------------------------+
|    Personal Assistant         |
|    Logic and Responses        |
+-------------------------------+
|
| 4. Actions and Responses
|
+-------------------------------+
|    Desktop Application         |
|  (e.g. web browser, calendar) |
+-------------------------------+
```

Explanation of the components:

1.   User's voice input: The user speaks to the personal desktop voice assistant, which captures the audio signal.
2.   Speech Recognition API: The audio signal is then sent to a Speech Recognition API, which converts it into text.
3.   Natural Language Understanding API: The text output from the Speech Recognition API is then sent to a Natural Language Understanding API, which analyzes the text and extracts intents and entities.
4.   Personal Assistant Logic and Responses: Based on the intents and entities, the personal assistant logic and responses are generated. These include actions and responses that the assistant can take, such as opening a web browser or scheduling a meeting.
5.   Desktop Application: Finally, the personal assistant interacts with the desktop application to perform the required actions or provide the user with the requested information.

VII. IMPLEMENTATION

The implementation of Personal Desktop Voice Assistant may involve several key steps, including:

- The **SpeechRecognition library** allows Python to access audio from your system's microphone, transcribe the audio, and save it.

- Google's **text-to-speech package, gTTS** converts your audio questions to text. The response from the look-up function that you write for fetching answer to the question is converted to an audio phrase by gTTS. This package interfaces with Google Translate's API.

- **Playsound package** is used to give voice to the answer. Playsound allows Python to play MP3 files.

- **Web browser package** provides a high-level interface that allows displaying Web-based pages to users. Selenium is another option for displaying web pages. However, for using this you need to install and provide the browser-specific web driver.

- **Wikipedia** is used to fetch a variety of information from the Wikipedia website.

- **Wolfram|Alpha** is a computational knowledge engine or answer engine that can compute mathematical questions using Wolfram's knowledge base and AI technology. You need to fetch the API to use thispackage.

VIII. ADVANTAGES

Some advantages of Personal Desktop Voice Assistant include:
- **Make up private conversations:**

With voice assistants, marketers have a new opportunity to start conversations in a more personalised way. With voice assistants, users typically express exactly what they want and are thinking. As a result, the channel enables marketers to respond with what they require and then keep in touch with customers to provide a customised experience.

- **Reach a large number of users at once:**

Marketers can reach several consumers in a single home thanks to voice assistants. Because each of these customers has different brand preferences, product interests, and music playlists, they all make distinctive purchasing judgements. One voice assistant allows marketers to collect more data and sell through a single targeted campaign, resulting in better outcomes.

- **Beyond the typical tools**

The fact that voice assistants are gaining popularity outside of our homes and mobile devices is an additional benefit. They are becoming more prevalent in our automobiles, smart Televisions, wearable technology, and home appliances. They offer fresh chances to accomplish even more goals while also giving current clients more value.

- **Achieve elusive prospects**

You have a better chance of reaching your target demographic by marketing via a virtual assistant on smartphones or smart speakers. It provides you with an alternative to doing so to the Internet and mobile. Voicebot.ai estimates that 87.7 million American people currently use smart speakers, an increase of 32 percent from January 2019 and an increase of 85 percent.

IX. DISADVANTAGES

Some potential disadvantages Personal Desktop Voice Assistant of include:

- **Concerns about data security**

Despite the fact that people are using voice assistants more frequently, there is still a lot of worry about the information these devices collect and the businesses that create the apps that run on them. Customers are concerned about the data's storage methods, viewers, and eventual disposition. If marketers don't handle these data and privacy issues, they won't be able to access these prospects or their data.

- **Disconnected exchange**

Another drawback is that compared to other platforms, voice assistants as a channel offer fewer enriching interactions. The choices include visual interactions versus speech material alone, which generally means recycling current content. This might make some of the more significant interactions that marketers can have elsewhere less effective.

- **Reliance on gadget manufacturers**

You are at the mercy of the manufacturers of devices, such as the corporations that build wearable tech, cars, and appliances, as a marketer. Before getting started, you should carefully investigate the device manufacturers you wish to collaborate with to achieve long-term success.

- **Investing in voice-activated apps and skill sets**

The cost of creating the voice app for this channel can be high. Building an internal skill set tailored towards the intricacies of voice assistants may take a lot of time if you participate in this channel. Consequently, it's crucial to weigh the advantages and disadvantages of using voice assistant channels.

X. CONCLUSION

In conclusion, personal desktop voice assistants are becoming increasingly popular as more people seek convenience and efficiency in their daily tasks. With the development of advanced natural language processing and machine learning technologies, these assistants can understand and respond to human queries in a more intuitive and human-like manner.

Personal desktop voice assistants have the potential to revolutionize the way we interact with our computers and devices, making it easier to navigate and access information. They can also enhance productivity and provide entertainment options, such as playing music or reading the news.

While personal desktop voice assistants have many benefits, there are also some concerns around privacy and security. Users need to be aware of the data that is being collected and how it is being used to ensure their personal information is protected.

Overall, personal desktop voice assistants have the potential to greatly enhance our daily lives and streamline our interactions with technology. As technology continues to advance, it will be interesting to see how these assistants evolve and improve in the years to come..

REFERENCES

[1]Palleti, V. (2021). SpeechRecognition. PyPI.

[2] Upadhyay, A. (2021). PyDictionary. PyPI.

[3] Sweigart, A. (2021). PyAutoGUI. PyPI.

[4] Yadav, M. (2021). pyttsx3. PyPI.

[5] Nallamothu, M., & Mukkamala, R. (2019). A Study on Speech Recognition and Desktop Assistant using Python. Journal of Emerging Technologies and Innovative Research, 6(4), 521-526.

[6] Mukherjee, A. (2019). Survey on Virtual Assistant (Google Assistant, Siri, Cortana, Alexa). 4th International Symposium SIRS 2018, Revised Selected Papers, 65-74.

[7] Haseeb, A., & Rizvi, S. A. R. (2022). A Review of Voice Assistant and Its Impact on the Healthcare Industry. International Research Journal of Engineering and Technology, 9(1), 299-303.

[8] Oracle. (2018). Advantages and Disadvantages of Voice Assistants for Marketers. Oracle Marketing Cloud Blog.

[9] Muneer, S., & Khan, F. A. (2022). Artificial Intelligence and Its Applications in Healthcare. International Research Journal of Modernization in Engineering, Technology and Science, 1(4), 23-27.

[10] Tripathy, S. (2020). Build Your Own Desktop Voice Assistant in Python. Analytics Vidhya.

[11] Kumar, A., Kumar, A., & Kumar, P. (2021). Development of Voice Assistant using Python. International Journal of Research in Engineering and Science, 10(2), 43-50.

[12] GitHub. (n.d.). Topics - desktop-assistant.