VIRTUAL DESKTOP ASSISTANT

Ms. Usha Dhankar[*1], Tanya Sharma[*2], Harshita Gupta[*3], Lovanya Bajaj[*4]

[*1,2,3,4]Guru Gobind Singh Indraprastha University, Department Of Computer Science And Engineering,

HMR Institute Of Technology & Management, New Delhi, Delhi, India.

ABSTRACT

An Intelligent Desktop Assistant (IDA) or AI-based Assistant (AIA) is typically a software program that executes some services or functions on private and direct command. Generally, in order to refer to some virtual assistants that can be accessed specifically through online chat, the term "chatbot" is employed. In the current date and time, there exists some online chatbots, as mentioned earlier, for the sole purpose of amusement and entertainment.

Several Virtual Assistants are developed enough to understand, analyse and interpret human speech. These assistants are also programmed to respond appropriately employing automated voices. Clients may question the desktop assistants or command them to automate media playback on their devices. These assistants allow the users to manipulate several other basic responsibilities such as calendars, e-mails and to-do lists through simple voice commands

Keywords: Artificial Intelligence, Machine Learning Models, Speech To Text Models, Text Analysis, Text Translation, Natural Language Processing.

I.   INTRODUCTION

The dependence of humans on others humans for several specifics tasks and services will soon cease. The technical advancement ensures that the need of seeking help of other human for various different specialization will be replaced by dependence on much better, reliable, efficient and secure technology for their regular requirements or other needs. Technology has become an integral part of human life. These devices allow every task ranging from basic maths calculations to multilayered programs making human work easy.

Desktop Assistants are slowly becoming and basic necessity for all the devices. It makes the execution and implementation of basic programs of specific requirements quicker the manual commands. Not just automated services, it also makes giving commands faster as the computers are equipped to translate human commands into machine language much faster than a human themselves.

One of the main unique features of a virtual desktop assistant that makes it extremely quick and easy to access is voice command. These voice commands are made possible with the help of voice recognition technologies. The basic concept of voice recognition in itself is a quite intricate procedure involving technologies like machine learning and neural networks. A neural network is formulated with vectors for each and every syllable and letter once the audio input is processed. This is termed as a dataset. When an input is received, the program computes different syllables on the basis of the highest match in comparison to these vectors.

Another necessity for these digital assistants is the mobile working culture where cars are quickly turning into mobile offices. This voice recognition feature will not only make the commands convenient; it will also prove to be a great safety measure.

Figure 1: AI assistant

Although, the indirect yield for the owners will be high, converging projects in the Virtual desktop Assistant industry such as VoiceTel, Mitel Networks, VoxSurf, WildFire, VoiceGenie and Conita are being offered by some establishments, while some provide solutions for the Mobile Carrier industry.

II.   LITERATURE REVIEW

A digitalized effort to initiate execution through a voice directed interface governing a subset of objects. The notebook memory stores a set of objects. Whereas from various fixed items, a subset is chosen; all of which possess an object type. This object type associates at least one taggable field and possesses a corresponding value. A spoken command is obtained from a person. it includes an instruction, a choice of an object type and a taggable field along with a transaction charge. In response to the speech, at least one item is retrieved from the gadget set, an item of the type selected by the user and whose price within the tagging area selection corresponds to the fee for the taggable field obtained from the user on which the command is executed. item. The object contains textual content that is converted to voice output [1]. They envisioned that one day computers would recognize natural language and calculate what we need, when and where we need it, and proactively take over for us. However, speech recognition and machine learning have continued to improve, and established records served through packages and content providers have emerged. We agree with this as computer systems are getting smaller and more ubiquitous [e.g. wearable devices and the Internet of Things (IoT) [2]. The recognizer is designed to change the verbal articulation from the individual to an alternative method of data (e.g. text). A single hand-held colleague containing a voice recognizer and a distinctive dialect processor is described. This piece of data can be a schedule for the day, data in an individual's diary, or data from an individual's address book, such as a phone number [3]. The most well-known use of the iPhone is "SIRI" which causes the end client to convey voice versatility to the end client in addition to responding to the client's voice charges. It is called Personal Assistant with voice recognition intelligence that takes the input of the client on a certain type of voice or content that processes it and returns the output in various structures such as the action to be performed or the item routed to the end client. In addition, this proposed framework can change the way of communication between the end client and mobile phones [4]. Open Data is currently collecting thoughts on the imaginative creation of the administration, mostly in the realms of government, life sciences, and smart enterprise. Be that as it may, he wants his application to move more for buyer administration, a web crawler for Open Data to realize what kind of information could help. This article introduces a voice colleague that uses Open Data as a teaching resource. It is highlighted by changing accuracy according to client criticism and obtaining unregistered information by client support. Furthermore, we demonstrate an application that helps with field work and confirms its viability [5]. The article presents a scheme of VPA applications and normal highlights and future patterns. The paper also proposes a tied-choice model in light of a quantitative assessment of the importance of requests and client availability. The Virtual Personal Assistant (VPA) is the coming age of carrier administration for portable clients. VPA is accepted as a smart advancement of the administration that caters to the regularly expanding demand of portability and networking professionals. VPA handles phone calls, handles individual exercises through a diary, 192 A. S. Tulshan and S. N. Dhage empowers the client to reach the business administrator through voice interfaces and includes all the elements of Unified Messaging. A virtual personal assistant (VPA) will allow the client to productively manage the growing interest in phone calls, messages, meetings and various exercises [6]. In any case, many people do not use them consistently. Past research has examined encounters with ongoing IPA clients. Using subjective techniques, we examine the experiences of rare clients: individuals who have tried IPA but choose not to use it consistently. It is clear that occasional clients share some of the consecutive client encounters, such as the displeasure of being trapped in a completely hands-off collaboration. Similarly, critical purposes of difference and previously unidentified concerns develop. Humanity IPA initiated correlations with human companions and compared their limitations. In particular, critical concerns have arisen regarding security, customization, permanence of information and directness. Based on these discoveries, we talk about the key difficulties, including outlining the ability to interrupt; re-examination of human likeness; issues of trust and ownership of information. A tendency towards these difficulties may lead to a wider use of IPA [7]. As virtual assistants become more intelligent and the biological community of IVA administration and gadgets expands, there is a growing need to understand the security and protection dangers of this growing innovation. Several late outbreaks contain a

notable vulnerability in IVA. Better demonstrative testing can reveal such vulnerabilities and prompt more reliable frameworks [8]. It allows target clients to connect to computers and web administrations with a wide range of usefulness in light of various web administrations and social media. There are four standard parts of the system; a voice recognition engine, a natural language processing engine, a conversational agent, and a content extraction engine. Current computer screen typing programs are not suitable for Internet access in terms of the basic assistance they provide for web content and the lack of participation in voice confirmation. Virtual right-hand programming open on the market is not particularly given to everyone and is not suitable for similar use. Some may face problems even now [9]. This paper presents the usability of four virtual voice and contextual text assistants (Google Assistant, Cortana, Siri, Alexa) [10]. Cortana can also read your messages, monitor your area, track your browsing history, check your contact list, keep an eye on your calendar, and set this information together to suggest valuable dates if you happen to enable it. You can also write your questions or requests if you don't want to get up indignantly. It is just a desktop based virtual assistant. [11] Siri: Siri has been an integral part of iOS since iOS 5 shipped in 2011. It started out with nuts and bolts, such as climate and inform, yet has expanded significantly since that point to help all outsiders. mix with MacOS. While Siri's jokes are incredible, the virtual assistant keeps getting more capable. Currently, you can request it to phone individuals, send messages, schedule meetings, send out applications and recreations, play music, answer questions, set up updates, and provide climate conjectures [12]. Google Assistant: Google Assistant (which now has consolidated capabilities from the more experienced Google as it is now being phased out) is unique in relation to Cortana and Siri. Virtual Assistant Survey: Google Assistant, Siri, Cortana, Alexa 193 A highly conversational VA is able to interpret basic vernacular and understand the importance of subtly complex requests such as, "What should we have for dinner? It can similarly see up to six unmistakable voices for couples and families, with each voice adapting to different diary events and inclinations, a great position for the Assistant to be amazing and flawless in a state where everyone uses a voice assistant on a unique device [13]. Alexa: While sharing various features similar to various VAs, Alexa is in a unique class of its own. Amazon's voice partner is not focused on portable or PC purposes, but instead on the standalone Amazon Echo speaker and Amazon Fire gadget set, with a stronger focus on whole house management and administration as opposed to PC errands [14].. Every business visionary, side hustler, and multitasking expert would love to have a virtual assistant right-hand man who could handle some of the boring daily tasks that accompany existing in an advanced age. As with any evolving innovation, it can be hard to isolate the build-up from the certainties in any case. Four notable players are vying for attention: Amazon (Alexa), Apple (Siri), (Google Assistant) and Microsoft (Cortana). I spent hours testing each of the four assistants by querying and charging fees that many business clients would use [15]. During the testing procedure, I noticed how the artificial intelligence reacts to me, and in addition, the intended users of the various components can think about, for example, the simplicity of the setup, the general ability to perceive my voice and the relevant understanding. Every mobile phone and computer available today have a brilliant right hand trapped inside, akin to a welcoming phantom - but how could they stand up to each other? While it may seem that Siri, Cortana and the mysterious Google Assistant are generally just an assortment of the same virtual partners, each of them has its own specific quirks, imperfections and characteristics. So, which one is best for clients? All things considered, this isn't a basic requirement to answer, as their point is that they're hard to look at without delving into their capabilities significantly. In this direction, we should start at this virtual right-hand connection [16]

III. PROBLEM STATEMENT

Create a virtual voice assistant so they can use the system to interact with new technology, control their devices, and use the technology for educational purposes. It is an ingenious technology for the visually impaired that provides them with auditory assistance, increasing the system's accessibility by using speech to text, text translation.

The system helps the consumer based on voice notes, i.e. the program acts on command issued by the user. The program will ask if something needs to be answered by the user because the user can't see the program running on the screen.

A visually impaired consumer will also be able to recognize himself as self-reliable.

The goal is to create a VDA that will be able to execute voice commands without mistakes.

The VDA will be able to understand the context correctly.

Several senior citizens face difficulty in operating new technology. Therefore, it would be useful to them too.

IV.   METHODOLOGY

Virtual Desktop Assistants employ Natural Language Processing to interpret the text version of the recognized voice command into machine understandable commands. Then these natural language or machine language commands are further processed by the software whenever the user asks the assistant to execute a task. These commands are then analysed by the software to generate an acceptable response. Some installable Python packages that are employed in the software coding of the virtual assistants are as follows:

Google-Text-to-Speech: As the name suggests, Text-To-Speech is typically used to translate any given text into speech of a chosen accent or nationality. In more complex terms, these engines create phonetical representation of the text, which are then converted to audible sound using the waveforms associated with the phonetics. Text-To -Speech has been advanced enough to provide the service in several different languages.

System calls: System Calls package employs a programming method which allows the computer program to ask for the service from the operating system kernel. These services can include hardware services like interactions with process scheduling, starting and running new processes or hard disk access. It proves to be a significant interface between the process and the OS.

Context extraction: Content Extraction package aids in automated extraction of structured information from unstructured or semi-structured machine-readable documents.

API calls: An API (Application Programming Interface) can be defined as an intermediate program or device that establishes communication between two different software products. In simple words, an API functions as a mediator, it sends user requests to the provider and the returns the response to the users.

Python Back-end: Python is generally employed for the entirety of the backend of the software. The context of the auditory command provided is then used to activate the appropriate response in the python backend, where the program decided if the required technology is a system call, API call or a context extraction. The formulated response is then further processed to complete the execution of the given auditory command.

Speech recognition: this program employs the use of Google's online speech-to-text services for speech recognition.

V.   FEATURES OF VDA

A. Tasks

A Task may be personal or work-related job that one would want completed. A Task can be a one-time job or a looped task. A looped task can either be scheduled to repeat at periodic intervals or it can be repeated as per a condition or a deadline set for it. For an instance, one would like to schedule in their appointment reminders for, say a month after their last appointment or they would like to send a PR mail every last Friday of the month to their clients. Repeated tasks are put in the task list one at a time. Once an occurrence of the task is marked as completed, the following occurrence appears in the list.

Clients can also use task requests. A task request refers to a request generated by a user to assign any specific task to another client or user. In simple word, one user asks another user to complete a task for them, while being completely capable of following its progress.

Task requests are required to maintain a task list, specifying the identities of those who originally generated the request, those who decided forward the request, those who decided to keep the updated copy, and those who finally accept the request. Virtual Desktop Assistant usually offers a Voice User interface for the MS Outlook Task Manager

B. Internet Applications

VPA allows users to obtain, modify or connect to the internet to access vast information and for various other purposes required to meet their daily task deadlines. Just with the use of a simple and convenient voice command, several tasks can be resolved like, hotel and flight reservations or visit planning, itineraries etc.

The combination of the vast internet with the fast mobile accessible culture of the handheld devices is now integrating with another new technology, creating several new possibilities of a voice web- A simple speech driven internet.

An internet-based network can be accessed through a computer only for now, but a voice interface is a way to achieve access to these networks through phones as well.

Figure 2: The task of virtual assistant

The developing Voice Webs (or Voice Portals) can be defined as auditory content distributed over the internet in the form of voice pages instead of web pages which can be easily accessed using a standard voice browser from any device. This innovation may also allow another business opportunity over E-Commerce as V-Commerce.

VI.   TECHNOLOGY USED

A virtual desktop Assistant should possess the ability to interpret the speech correctly, formulate acceptable responses and actions for the given instructions, and finally deliver appropriate results desired by the users.

Programming Language: One of the most popular and utilized programming languages for the time being is **Python**. Python programming is equipped with a treasure of frameworks, libraries and packages making implementation of technologies as complex as Natural Language Processing, Machine Learning and Speech Recognition much faster and easier. Python also allows simple integration of program code with the internet accessible APIs, namely, Spotify API, Google Cloud Speech-to-text API, Weather API, Maps API, IP Adress APIs, and several other facilities.

Graphical user Interface (GUI): Python also offers a variety of GUI packages including Tkinter and PyQT. These provide a visually attractive and interactive face to the Virtual desktop Assistant.

Machine Learning Frameworks: Machine Learning Frameworks are required for the training of Machine learning models for NLP and speech recognition jobs. Some of the most commonly used python libraries offering machine learning frameworks are Scikit-Learn, PyTorch, and TensorFlow.

Text-to-Speech Engines: Python also offers libraries for Text-to-Speech conversion service, namely, pyttsx3, gTTS, larynx, TensorFlowTTS etc. these libraries allow generation of human like speech from textual data either offline or online.

Natural Language Processing (NLP) Package: Some of the more famous NLP libraries offered in Python are Stanford CoreNLP, spaCy, and Natural Language Toolkit (NLTK). Some of the main tasks taken by these libraries can be semantic analysis, identification of named entities, key word tagging etc.

VII. REQUIREMENTS

Software Requirements:

Windows OS, MacOS, Linux, PyCharm, Visual Studio

Hardware Requirements:

4GB RAM, Microphones.

Other Requirements:

Strong Internet Connection

Functional requirements:

Some of the important features for a virtual desktop assistant to have been listed below:

- **Constraints:** Virtual desktop Assistants are required to be compatible with all the operating systems to make independent.
- **Usability**: A virtual desktop assistant should be easily accessible and easy to interact. It must have convenient activation calls.
- **Security:** Data security is another main concern for a VDA systems.
- **Response time**: Quick responses should be a must to increase convenience and comfort.
- **Accuracy**: Voice command recognition and interpretation must be accurate.
- **Multi-language support**: Multiple language interpretation and translation should be available.
- **Performing tasks:** The program should have the ability to perform major tasks as well as menial tasks such as setting reminders, setting alarms, sending emails, playing music, pausing music etc..
- **Natural language processing:** command interpretation should possible and accurate in any required language.
- **Wake word detection:** There ought to be some kind of an activation or waking call available like "Hello Assistant" or a maybe just a clap.

VIII. SYSTEM ARCHITECTURE

The system architecture of a program defines and represents the control flow among the various components of the system.

The system architecture for a Virtual Desktop Assistant encompasses various components that work with each other to complete several major as well as menial tasks and offer a variety of functionalities.

A Virtual Desktop Assistant majorly includes;

- A UI to interact,
- NLP Module for command interpretation,
- A knowledge base for retrieval and storage of information,
- A module to record execution of action while completing tasks,
- A layer for integration with external services,
- A Module that records personalization and Learning information to adapt,
- A security layer for data protection,
- Logging mechanisms to monitor usability, and
- Deployment and scalability measures.

It makes sure that the interaction between the assistant and the user remains smooth-sailing and comprehensive.
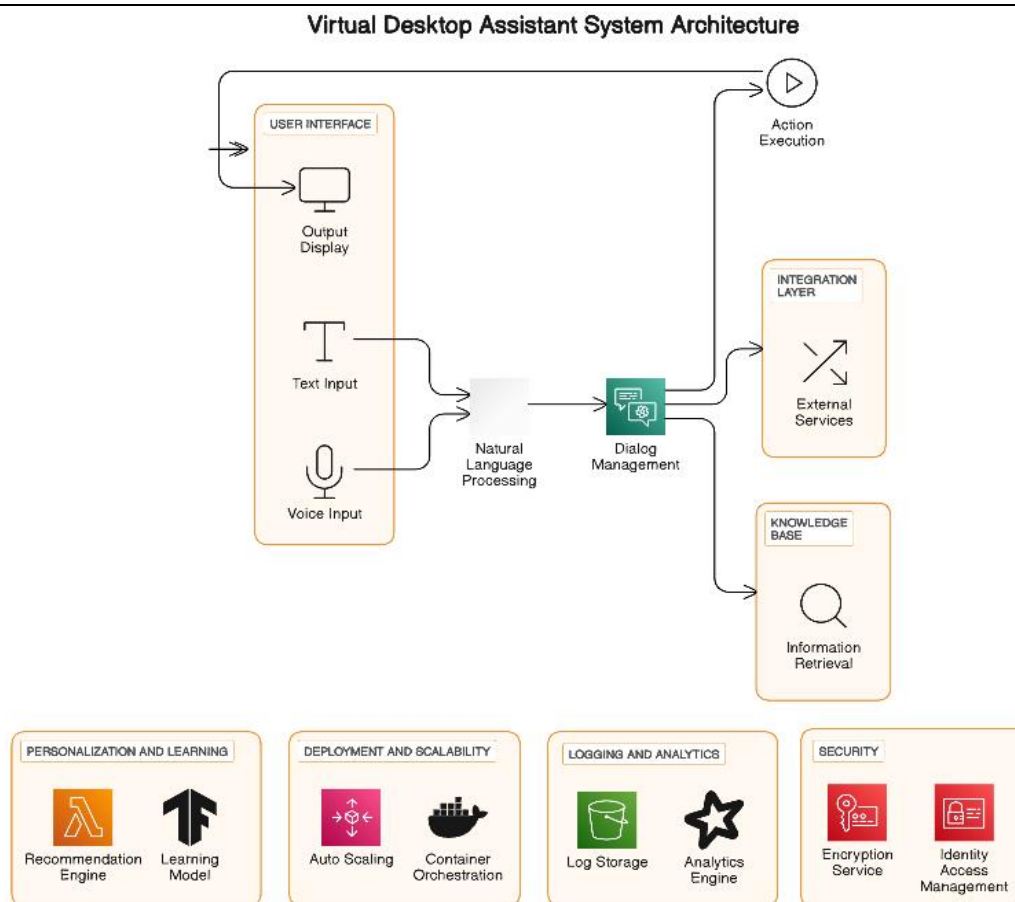
Figure 3: System Architecture

IX.   EXISTING AND PROPOSED SYSTEM

Existing Model

Majority of the existing system or programs make use of the neural networks for the purpose of speech recognition. Even though the accuracy levels are high, the complexity of neural networks raise issues in the efficiency department for the users.

Some the most used techniques can be listed as below:

1. NLP:

Natural Language Processing is subdivision of Artificial Intelligence which makes interaction among machine and human languages possible. It's mostly based on the idea of programming computers to process large blocks of data in natural languages. This segment of machine familiarizes the machine with several unique and different words in any language whilst allowing the machine to recognize and interpret those word in general speech.

2. MFCC (Mel-Frequency Cepstral Coefficients):

Mel-Frequency Cepstrum or MFC can be defined as a set of coefficients. It refers to the short-term energy spectrum of sound that is used to sense changes in speech that helps in the recognition of the several variables required in voice recognition.

3. Context-aware computing:

Context-aware computing can be categorized among the systems that possess the awareness of their physical surroundings granting them the ability to understand accented speech as well as mispronounced or contextually incorrect words.

Proposed speech-to-text model

This paper proposes the use of an intermediary software capable of auditory input into textual output but without the contextual awareness what so ever.

Text Analyzing:

- The textual output has no context for the machine.
- A block of algorithms gives meaning to this text to provide understandability to the machine.
- Once the text is machine understandable the virtual assistants will derive the direct command from the text.
- VPAs map functions and parameters to the syllables in the perceived textual command and provides instructions to the machine about the given tasks.

The main issue this project is trying to address is the low accuracy rate of the speech recognition software. This project is set top create a technology that will be capable of accurately recognizing any kind of accented speech even with the slightest of modulations with a great level of precision. This is technology is a must for the virtual assistant to be of any real use in today's time.

Simple voice recognition is of no real use when accented speech is encountered as the spoken words will differ a great deal from the actual non-accented pronunciation as in the machine's database.

This issue can be remedied through lip movement recognition.

For the majority of the words, even with a different accented pronunciation, the lip movement remains similar enough to help recognize the correct word. Thus, the lip movement recognition is another beneficial factor that can be added to the speech recognition software as an upgrade to the only voice recognition software.

X. WORKING PRINCIPLES

The main principles followed by general virtual assistants are listed below:

Natural Language Processing: In simple words, Natural Language Processing is a medium that allows us to communicate with machine using our own human languages that are also called natural languages.

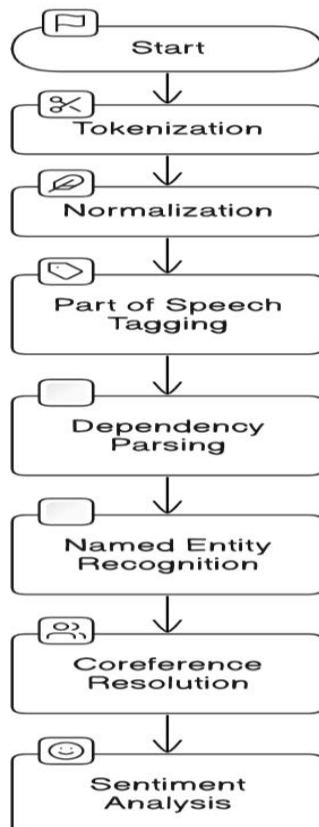The complex algorithm of NLP includes several steps:

Figure 4: Steps in NLP

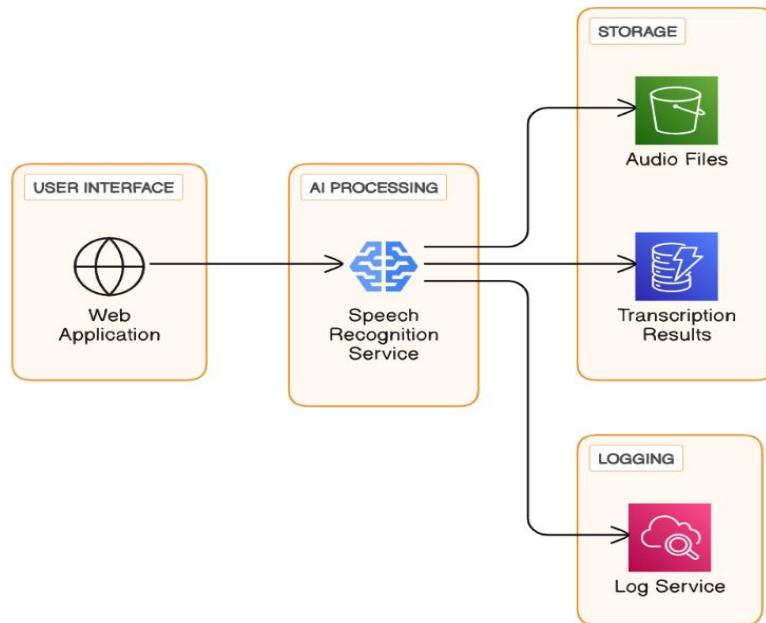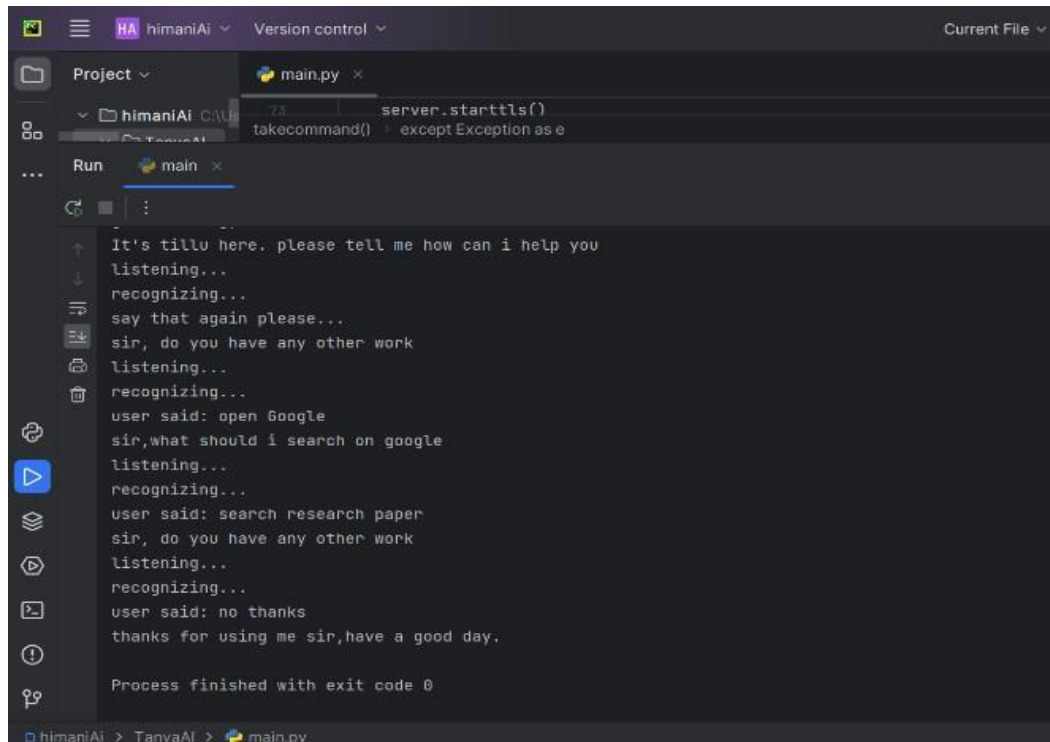**Automatic Speech Recognition:** For the quick interpretation of user commands.

Figure 5: Speech Recognition Architecture

Machine Learning: Automated study and storage of user preferences and behaviour is another benefit in the virtual assistant architecture is it allows the use of virtual assistant more efficient and convenient for the users.

• **Inter-system verbal exchange:** To recognize and receive required data from other packages and other software applications.

XI.   RESULT

Figure 6: Listening and recognizing
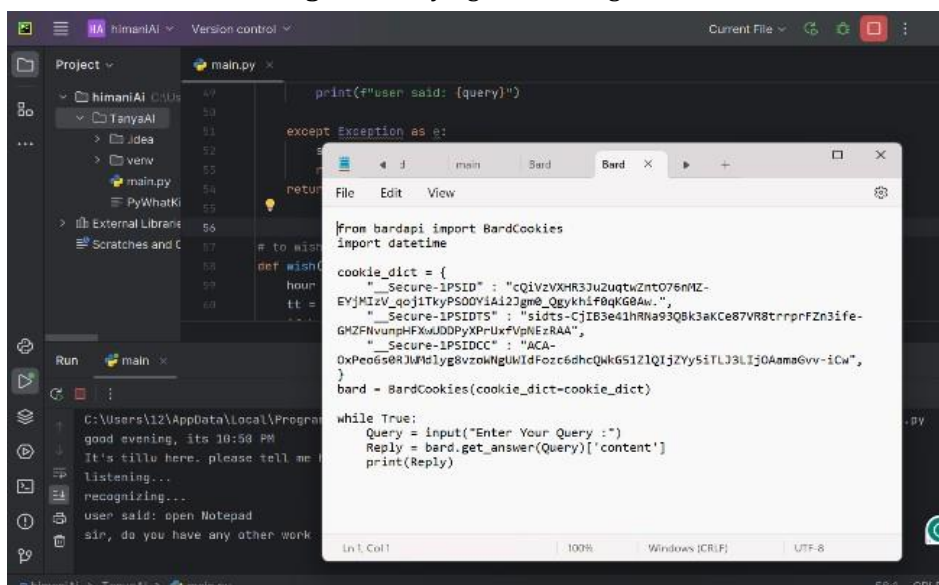
Figure 7: Playing music using VDA

Figure 8: Open Notepad using VDA

XII.   CONCLUSION

Virtual Desktop Assistants are increasing gaining popularity as more and more people are being introduced to the fast-paced folds of life.

Efficiency and convenience are becoming a required need for all of us. With all the advancement in technology, these assistants are popularly gaining attention providing quick responses, better utilization of time and faster task completion abilities.

In addition to normal people, these assistants prove to be of great value to the visually aided people, making the use of technology possible and efficient for them.

But, as it is said there is no end to the shortcomings, or, nothing is perfect.

Just like that, these assistants have their own challenges to overcome and issues to be resolved still. Starting with privacy concerns and ending with response time, these assistants are still in the improvement phase of their life cycles.

Therefore, any users should stay aware of the data that is processes by these assistants and that is used to build a personalized bot. Users should keep an eye on how this data is used and where it is being stored.

XIII. REFERENCES

[1] Gong, L.: San Francisco, CA (US) United States US 2003.01671.67A1 (12) Patent Application Publication c (10) Pub. No.: US 2003/0167167 A1 Gong (43) Pub.

[2] Date: 4 September 2003 for Intelligent Virtual Assistant Sarikaya, R.: The technology behind personal digital assistants. IEEE Signal Process. Mag.34, 67–81 (2017). https://doi.org/10.1109/msp.2016.2617341

[3] Tsiao, J.C.-S., Tong, P.P., Chao, D.Y.: NaturalLanguage Voice-Activated Personal Assistant, United States Patent (10), Patent No.: US 7,216,080 B2 (45), 8 May 2007 4. Siri, K., Patankar, A.J.: Personal assistant with voice recognition intelligence. Int. J. Eng. Res. Technol. 10(1), 416–419 (2017). ISSN 0974-3154

[4] Kawamura, T., Ohsuga, A.: Flower voice: virtual assistant for open data Elshafei, M.: Virtual personal assistant (VPA) for mobile users. Mitel Networks (2000– 2002)

[5] Chung, H., Iorga, M., Voas, J., Lee, S.: Alexa, can I trust you? In: 2017 IEEE Computer Security (2017)

[6] Cowan, B.R.: What can I help you with?: infrequent users' experiences of intelligent personal assistants. In: 2015 IEEE 10th International Conference on Industrial and Information Systems, ICIIS 2015, Sri Lanka (2015)

[7] Weeratunga, A.M., Jayawardana, S.A.U., Hasindu, P.M.A.K, Prashan, W.P.M., Thelijjagoda, S.: Project Nethra - an intelligent assistant for the visually disabled to interact with internet services. In: 2015 IEEE 10th International Conference on Industrial and Information Systems (2015)

[8] López, G., Quesada, L., Guerrero, L.A.: Alexa vs. Siri vs. Cortana vs. Google Assistant: a comparison of speech-based natural user interfaces. In: Nunes, I. (ed.) AHFE 2017. AISC, vol. 592, pp. 241–250. Springer, Cham (2018). https://doi.org/10.1007/978-3-319- 60366-7_23

[9] Zhao, Y., Li, J., Zhang, S., Chen, L., Gong, Y.: Domain and speaker adaptation for Cortana speech recognition. In: ICASSP

[10] Bellegarda, J.R.: Spoken language understanding for natural interaction: the Siri experience. In: Mariani, J., Rosset, S., Garnier-Rizet, M., Devillers, L. (eds.) Natural Interaction with Robots, Knowbots and Smartphones, pp. 3–14. Springer, New York (2014). https://doi.org/10.1007/978-1-4614-8280-2_1

[11] Google: Google Assistant. https://assisatnt.google.com

[12] Purington, A., Taft, J.G., Shannon, S., Bazarova, N.N., Taylor, S.H.: Alexa is my new BFF: social roles, user satisfaction, and personification of the Amazon echo. ACM, 6–11 May 2017. ISBN 978-1-4503-4656-6/17/05

[13] Lopez, G., Quesada, L., Guerrero, L.A.: Alexa vs Siri vs Cortana vs Google Assistant: a comparison of speech-based natural user interfaces. Conference Paper, January 2018

[14] Kepuska, V., Bohouta, G.: Next generation of virtual personal assistants (Microsoft Cortana, Apple Siri, Amazon Alexa, and Google Home). In: IEEE Conference (2018)