Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 15, Issue. 1, No.11: 2024
ISSN : **1906-9685**

Voice Based Virtual Personal Assistant

Dr. N.Thirupathi Rao, M.V.Sushriya, M. Navya Manaswi,
.V.N.S.Aishwarya,P Vikas, P Surya Teja
Department of Computer Science and Engineering, Vignan's Institute of Information Technology (A), Visakhapatnam, AP, India.

ABSTRACT

The project aims to develop a personalized virtual assistant tailored for Windows and Linux systems, drawing inspiration from prominent virtual assistants like Cortana and Siri. Named Leo, it is designed to provide users with a user-friendly interface, facilitating various tasks through clear and well-defined commands. It enables interaction through both voice commands and traditional keyboard input methods and as a personal aide, it assists users with everyday activities, including engaging in casual conversations, conducting searches on platforms such as Google or Yahoo, finding videos, playing music, providing live weather updates, defining words, fetching details about medicines, and reminding users of scheduled events and tasks. Utilizing voice input captured via the microphone, it processes spoken language into computer-readable data, offering solutions and answers to user queries. Leveraging internet connectivity, it swiftly retrieves requested information, enhancing users knowledge , providing prompt responses and delivers responses through voice commands and also by displaying text on the screen for added clarity. The primary objective of this virtual assistant is to augment users intelligence by delivering instant and calculated results.

Keywords: Voice-based interaction , Virtual assistant , Natural language processing , Speech recognition , Machine learning , Personalization , Interactive feedback

1.INTRODUCTION

Within the fast-paced world of advancing technologies , the emergence of voice-controlled virtual helpers has upended how people connect with machines. Imagine a world where tasks are accomplished with a simple command, where information is at your fingertips just by speaking. This project delves into the realm of these intelligent companions, exploring their emergence as indispensable tools in our daily lives. From organizing schedules to controlling smart devices, voice-based virtual personal assistants have seamlessly integrated into our routines, offering convenience

http://doi.org/10.36893/JNAO.2024.V15I1N01.334-348

and efficiency like never before. We invite you to come along on our endeavor to disentangle the nuanced inner workings of these state-of-the-art tools and to imagine the limitless opportunities they portend for computing in the years yet to be.

1.1. About Voice-Based Virtual Desktop Assistant

In our increasingly technology-driven world, the demand for user-friendly interfaces facilitating interaction between humans and computers is on the rise. Voice-based virtual desktop assistants epitomize a watershed moment for the field, allowing users to engage with their machines hands-free through a streamlined means of communication without physical interaction.

a. Voice Recognition: Leveraging advanced speech recognition algorithms, the virtual desktop assistant processes spoken commands and queries, enabling users to perform tasks effortlessly through natural language interaction.

b. Personalized Assistance: Tailored for Windows systems, the virtual assistant, named Leo, draws inspiration from renowned counterparts like Cortana and Siri. Leo is crafted to provide a user-friendly interface, allowing users to execute various tasks using clear and defined voice commands or traditional keyboard input methods.

c. Extensive Functionality: Leo serves as a personal aide, supporting users with a wide array of everyday activities. From engaging in casual conversation to conducting searches on popular platforms, playing music, providing weather updates, and setting reminders, Leo enhances productivity and convenience in users' computing experiences.

d. Intelligent Response: Utilizing voice input captured via the microphone, Leo processes spoken language into computer-readable data, offering solutions and answers to user queries with swift accuracy. Additionally, Leo employs internet connectivity to retrieve requested information promptly, enriching users' knowledge and providing timely responses.

e. Enhanced Accessibility: Operating predominantly through voice commands, Leo not only delivers spoken responses but also displays relevant information on the screen for added clarity. This feature ensures accessibility for users with visual impairments or those who prefer visual reinforcement.

http://doi.org/10.36893/JNAO.2024.V15I1N01.334-348

Fig.1.Diagram depicting how a voice based virtual assistant work

2.Literature Survey

The rise of voice interaction has revolutionized human-computer interfaces. Virtual desktop assistants (VDAs) leveraging voice commands offer a hands-free, intuitive approach to interacting with computers. This review explores the current landscape of VDAs, focusing on:

Existing Solutions:

Briefly discuss prominent commercial VDAs like Google Assistant and Amazon Alexa, highlighting their functionalities and limitations, particularly in the desktop environment. Consider referencing a research paper analyzing these limitations in a desktop context, such as:

A. Pasztecka et al., "Usability Evaluation of Intelligent Virtual Assistants on Desktop Computers: https://www.researchgate.net/publication/348400306_Usability_Evaluation_of_Artificial_Intelligence-Based_Voice_Assistants_The_Case_of_Amazon_Alexa"

Core Technologies:

In discussing the core technologies powering voice-directed applications, one must outline how Automatic Speech Recognition allows for comprehending vocal utterances while Natural Language Processing facilitates interpreting instructions conveyed through spoken commands.

Jurafsky, Daniel, and James H. Martin. Speech and Language Processing (3rd Edition). Pearson, 2019: This is a textbook reference

Benefits and Applications:

By highlighting how virtual desktop assistants can help boost efficiency through higher output, accommodate individuals with impairments, and allow for enhanced juggling of multiple tasks simultaneously, one stresses the benefits these systems offer.

M. Khalil et al., "The Impact of Virtual Desktop Assistants on User Productivity: https://www.linkedin.com/pulse/impact-virtual-assistance-workplace-productivity-era-k-abdullah-bz5xf/

Challenges and Future Directions:

Acknowledge challenges like limited functionality compared to traditional interfaces, potential privacy concerns, and language barriers.

S. Fahl et al., "Voice Assistants: Alexa, Siri, Google Assistant and How They Can Spy on You: https://www.cnbc.com/2018/05/13/are-siri-alexa-and-google-assistant-spying-on-me.html

3.Existing System

Most of the existing projects have primarily relied on neural networks for speech recognition, yielding moderate accuracy levels. While such systems may propose intriguing theories, deficiencies in implementation and effective function often relegate them to theoretical idealism rather than viable solutions for tangible problems.

Innovative Adaptability:

At the forefront of technological advancements lies innovative adaptability. Context-aware computing represents a paradigm shift in system functionality. With the capability to comprehend words spoken in diverse accents and rectify mispronunciations, they significantly elevate accuracy and performance.

Advanced Sound Analysis:

Venturing into technical intricacies, advanced sound analysis techniques come to the forefront. One such technique is Mel-Frequency Cepstral Coefficients (MFCC), representing a significant breakthrough. This sophisticated approach enables systems to detect subtle variations crucial for voice recognition, thus enhancing accuracy in identifying diverse voice parameters.

Language Understanding:

Another critical aspect of cutting-edge technology is language understanding. This field delves deeper into comprehending human language. By analyzing and interpreting natural language data, systems can grasp the nuances of spoken words with remarkable precision and fluency.

4.Proposed System

The proposed system aims to improve speech-to-text software by enhancing its accuracy and usability for various accents and speech patterns. It involves the development of software that converts spoken words into text with high precision, making it suitable for everyday use in virtual personal assistants (VPAs) like Siri.

Key Components:

Speech-to-Text Conversion:

Software designed to accurately convert spoken words into text format.

Focus on improving accuracy even with different accents and speech variations.

Usability Enhancement:

Aim to make the software user-friendly and accessible for daily tasks.

Potential integration with virtual personal assistants for seamless interaction.

Project Development:

The development process involves using neural networks for voice recognition and machine learning algorithms for lip movement detection. By combining these technologies, the software aims to achieve higher accuracy in converting speech to text, even for individuals with diverse accents.

4.1.Advantages of the proposed system:

The proposed voice-based virtual desktop assistant offers numerous advantages that enhances :

- Accessibility

- User Feedback
- Convenience
- Improved User Experience
- Natural Language Understanding
- Personalization
- Continuous Improvement

5.Methodology

The methodology section will delve into the development process of your voice-based virtual desktop assistant (VDA). Here, you'll dissect the system architecture, outlining the various modules that work together to understand user requests and execute actions. These modules will likely include:

Speech recognition: This module converts spoken commands into text.

Natural Language Processing (NLP): This module interprets the converted text, extracting user intent and relevant information.

Knowledge base and information retrieval: This component houses the information the VDA utilizes to respond to user queries. It could be a local database, external API integration, or a combination of both.

Action execution: This module translates user intent into concrete actions, such as controlling system functions, launching applications, or interacting with web services.

Text-to-speech (TTS): This module converts generated text into audio output.

http://doi.org/10.36893/JNAO.2024.V15I1N01.334-348

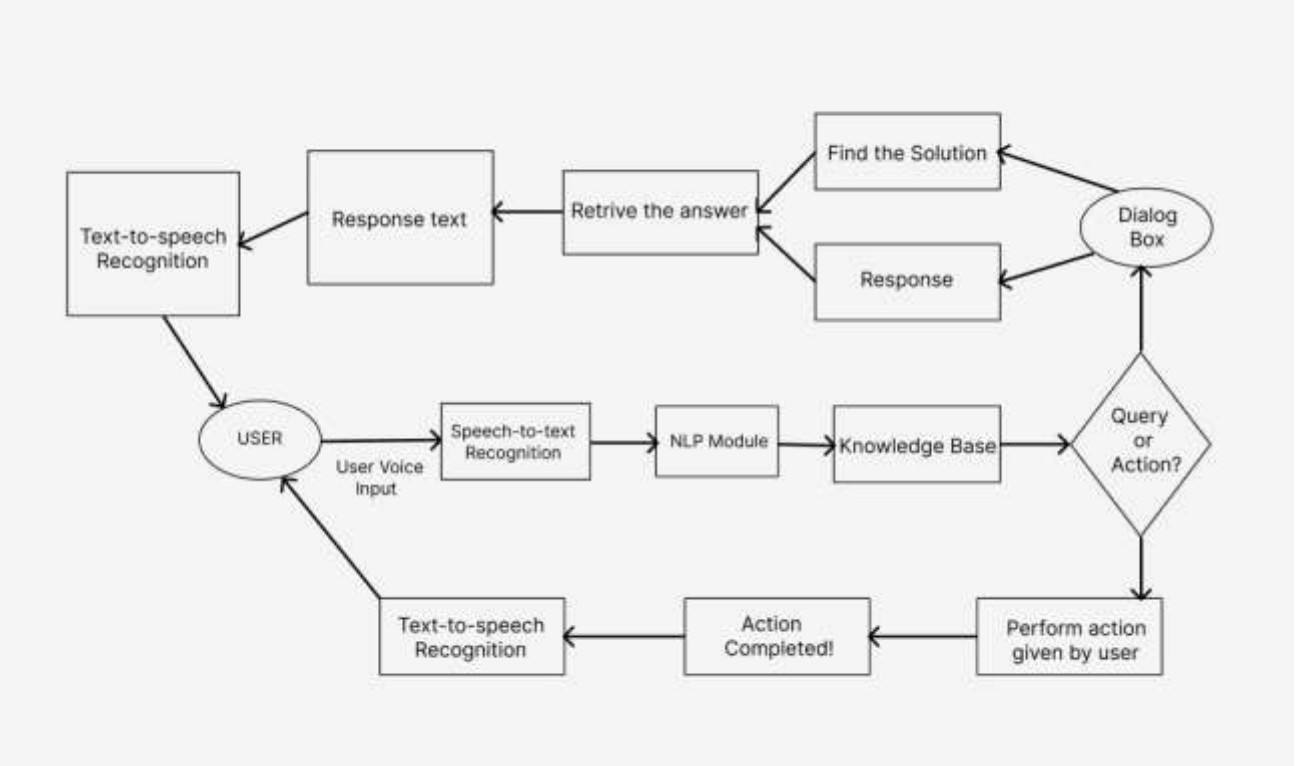

Fig.2.Flow Diagram of the model

5.1.Architecture Diagram of Voice Based Virtual Personal Assistant

The voice-based desktop assistant architecture includes modules for speech recognition, natural language processing (NLP), and task execution. Speech recognition converts spoken commands to text, while NLP infers user intent for generating responses. Task execution interfaces with applications to perform actions. Accessibility features, like audio feedback, aid users with impairments. This architecture enables efficient task completion through voice commands.

The Architectural Diagram of Voice Based Virtual Personal Assistant is depicted in the fig. diagram below.

http://doi.org/10.36893/JNAO.2024.V15I1N01.334-348

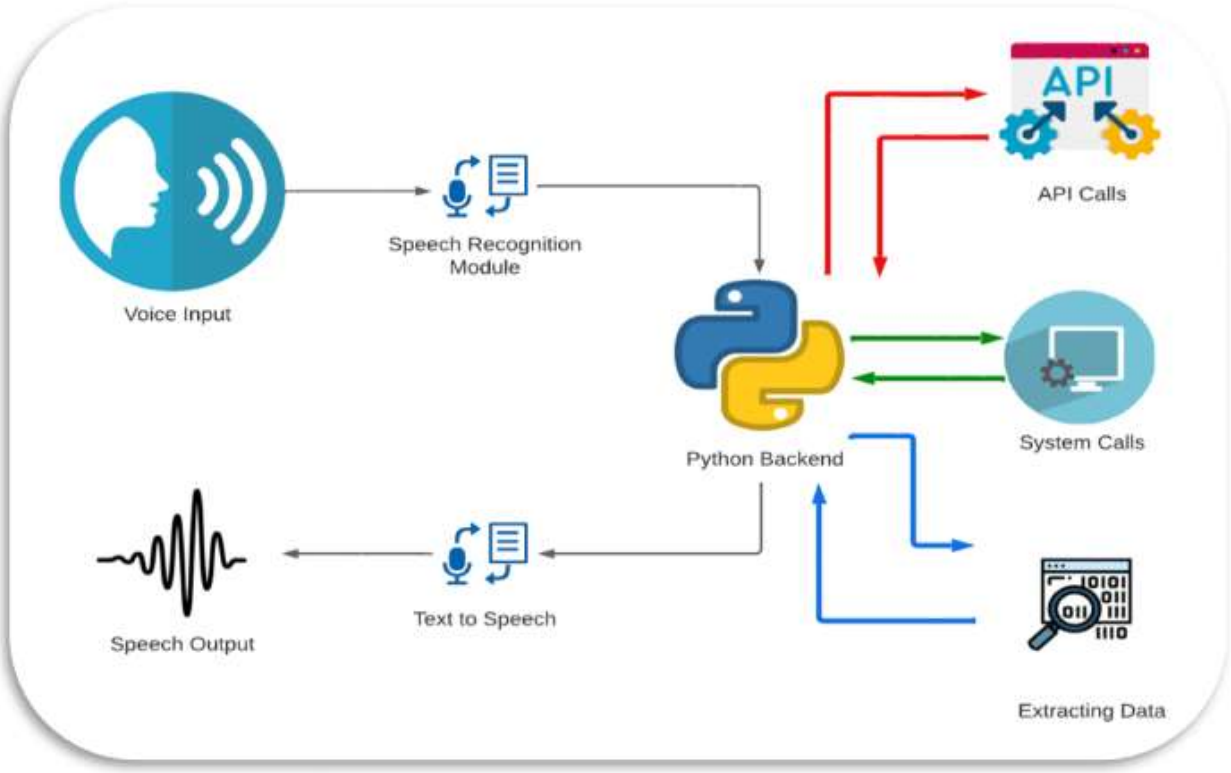

Fig.3. Detailed workflow of the model

Among the fundamental aspects that comprise the design are several noteworthy elements of the construction plan.

- **User**: This is the person interacting with the virtual assistant.
- **Speech Recognition Module**: This module converts the spoken voice input from the user into text.
- **Text**: This is the representation of the user's spoken input as text.
- **Python Backend:** This is the program that extracts data from the text input, and uses system calls to carry out actions.
- **System Calls:** These are requests made by the Python Backend to the operating system to perform tasks.
- **External API:** These are APIs that provide data or services to the virtual assistant.
- **Database (DB):** This is a storage location for data that can be accessed by the virtual assistant.
- **Text to Speech Module:** This module converts the text response generated by the Python Backend into spoken language.
- **Output:** This is the spoken response that the virtual assistant delivers to the user.

5.2.Algorithm

For implementing face recognition , we used an algorithm called LBPH algorithm.

FULL FORM: LBPH stands for **local binary pattern histogram.**

http://doi.org/10.36893/JNAO.2024.V15I1N01.334-348

There are 2 types of data that we are using- test data and train data.

Factors affecting LBPH: The LBPH algo mainly depends on 4 factors. they are:

grid x- number of cells in horizontal direction . usually set to 8

grid y- number of cells in vertical direction . usually set to 8

neighbors-number of sample points to build local binary values

radius- used to build the local binary pattern which is usually set to 1.

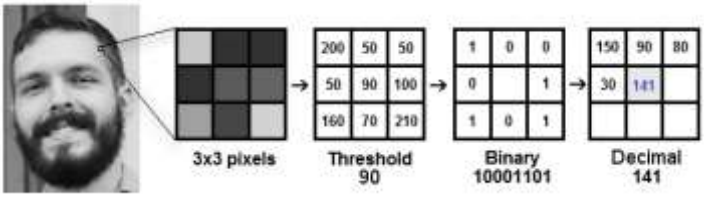

Fig.4. The working of LBPH algorithm

Input: In the LBPH algorithm we first take the training data . The training data is taken in a grayscale image. now the image is divided in 3*3 pixels

Assigning Values: The center pixel is called the threshold and it has 8 neighbors . The threshold value depends on the neighboring values. if the neighboring value is less than the threshold value then we assign ‘0’ to that value and if the neighboring value is greater than or equal to threshold then assign ‘1’.

Calculating Binary Values: now we calculate the binary values for each position by concatenating the binary values. After this we will set the obtained binary value to the decimal value and assign it to the center pixel . like this we will calculate for the whole image. Then we will obtain the histogram for the taken image. like this we will get the local binary pattern value for each image .

Test Data: Test data is given at the time of program execution where it comes with the inbuilt local binary pattern values.

Detection of Faces: Now we take the training data and test data and subtract the local binary pattern values . This subtracted value is called the euclidean value . if the euclidean value is ‘0’ then that image is the perfect match and the output is given.

Output: After the output is given then the system will see whether that particular person will get access to the system or not.

LBPH is renowned as one of the most straightforward face recognition algorithms. It is predominantly favored for its ability to identify side faces or slightly blurry images with remarkable

http://doi.org/10.36893/JNAO.2024.V15I1N01.334-348

accuracy. Implemented through OpenCV, LBPH consistently delivers outstanding results, distinguishing it as a preferred choice among various algorithms.

6.Results and Discussion

Output 1:

The news: Upon user inquiry, Leo promptly delivers the most recent news in both text and speech formats without delay.The news appears right on the terminal screen, without being sent to any external websites.

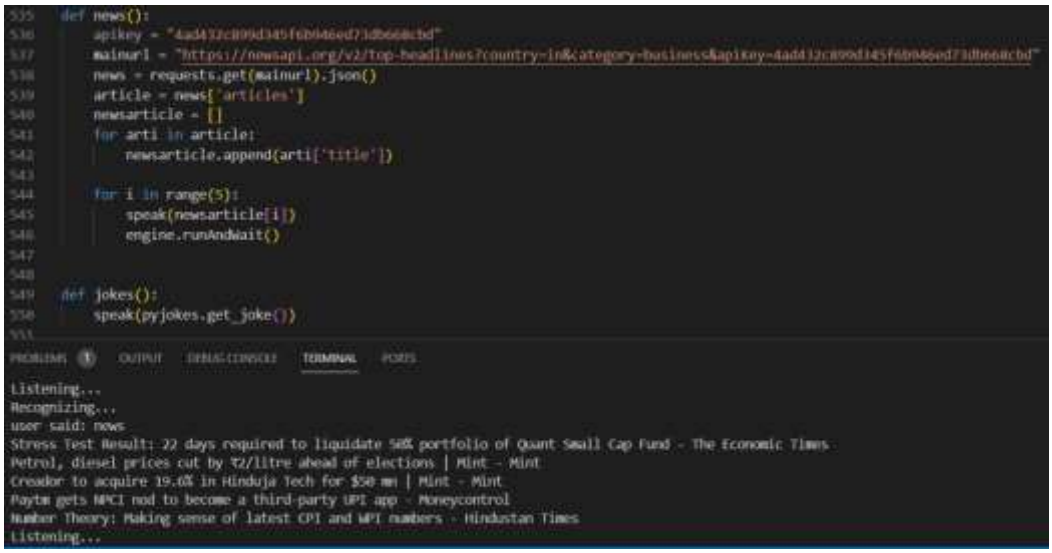

Fig.5.:The news output is shown as above

Output-2

Here, the user asked leo to open youtube & start a cybersecurity video. This screen also helps us understand that we can also google our queries and a few more actions.

http://doi.org/10.36893/JNAO.2024.V15I1N01.334-348

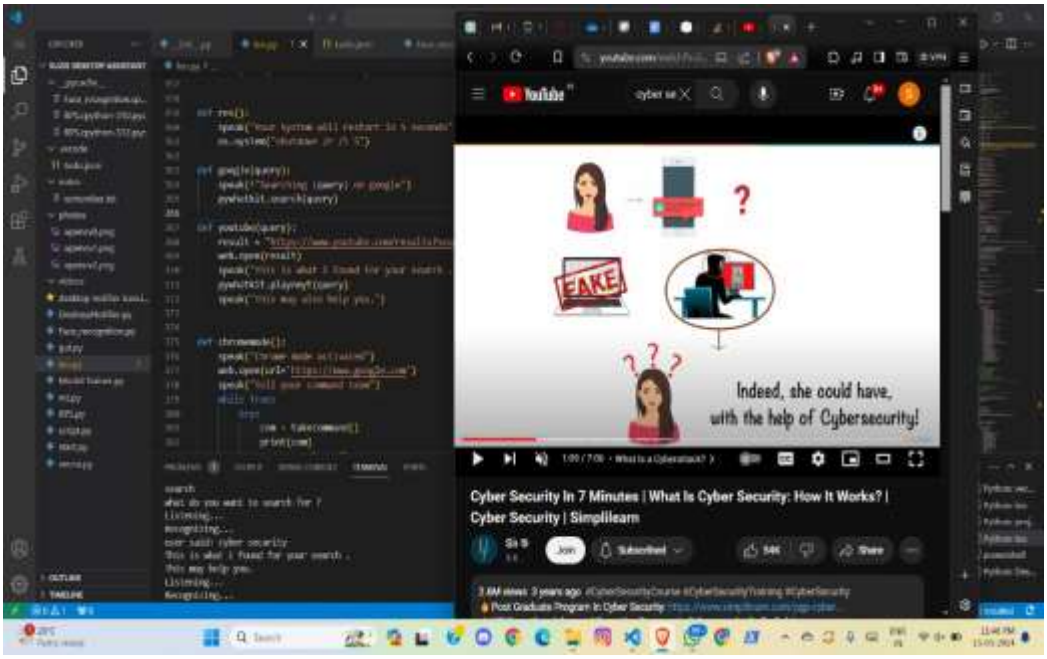

Fig.6.: The output shows the youtube window

Output-3:

Capturing a photo: Using opencv, we are able to click a picture using voice command.It takes just a few seconds to capture it and then save it at a location too as we already can see three pictures saved here.

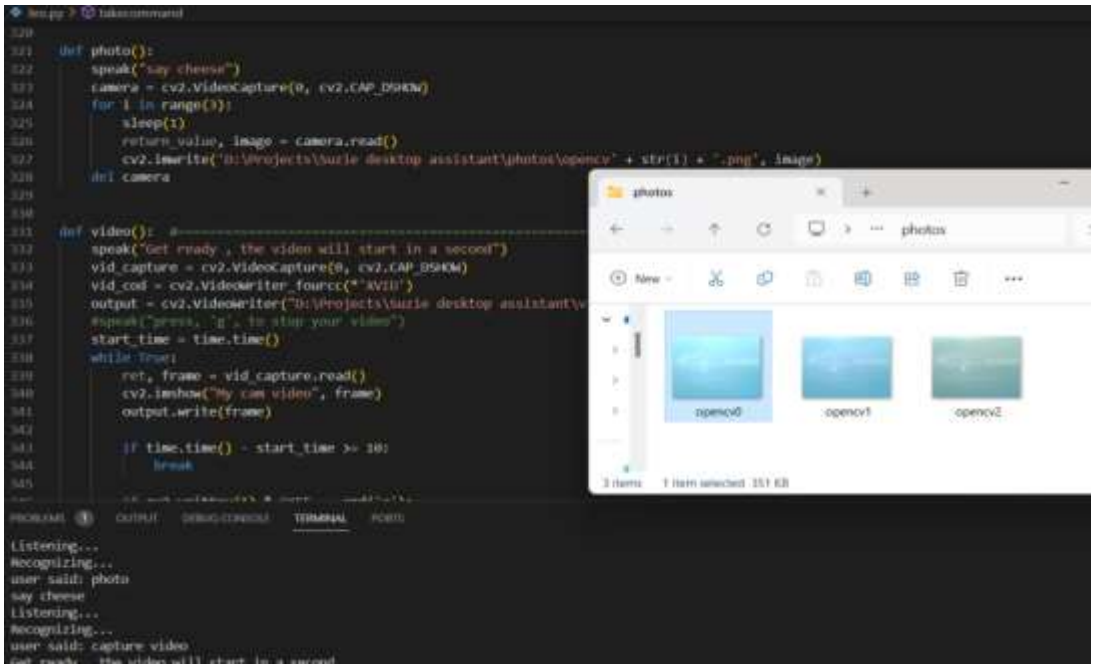

Fig-7.: Capturing a photo

Output-4:

http://doi.org/10.36893/JNAO.2024.V15I1N01.334-348

In this, we are also able to record a video by opening the camera and enabling the cam in our system. We can record till how much time we want it to record.

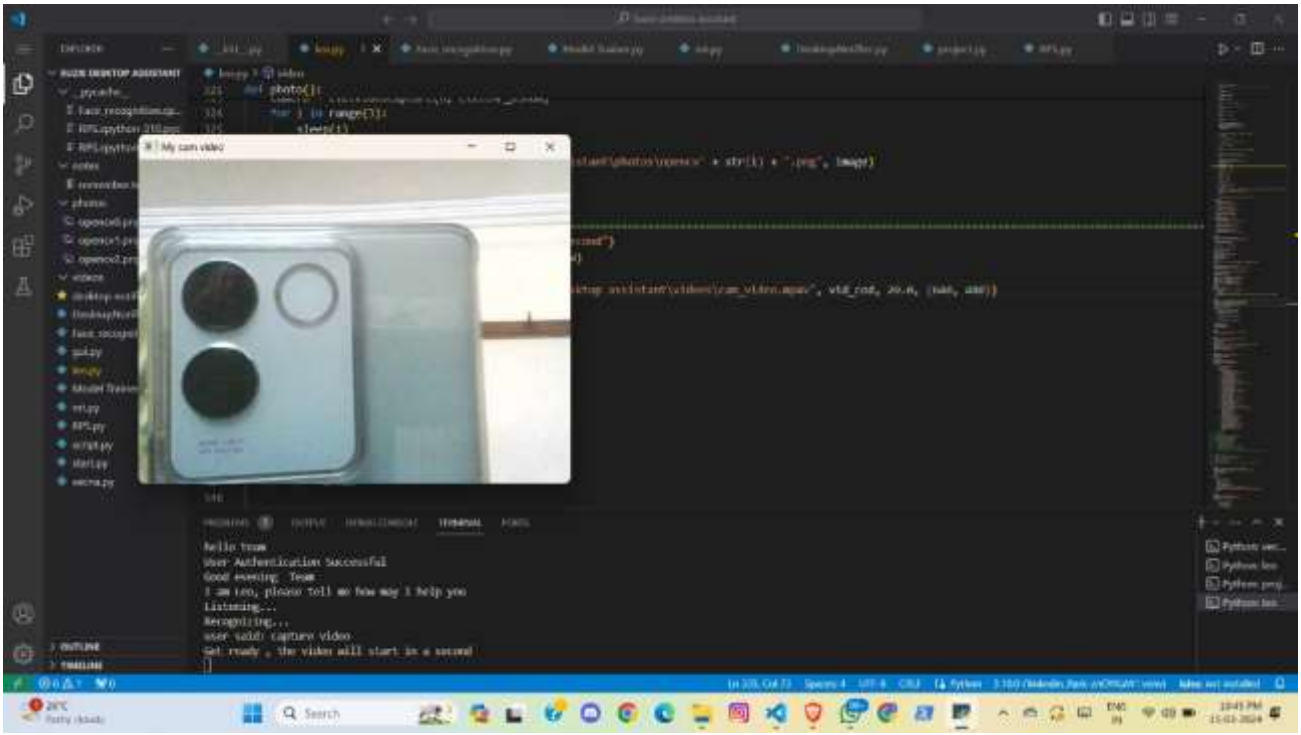

Fig.8.: Output shows the video is being captured

Output.5:

In youtube, we can not only start a video, pause, play, fast forward, but we can also download videos into our local systems. Any video which allows download can be done through this application.

Fig-8: Output shows a downloaded video from youtube

http://doi.org/10.36893/JNAO.2024.V15I1N01.334-348

6.4.Applications of Voice-Based Virtual Desktop Assistant

Implementing the voice-based virtual desktop assistant can bring about a multitude of benefits and applications:

I. **Enhanced Communication Efficiency:**

Through enabling voice interaction with computers and devices, the virtual assistant is able to streamline communication workflows while lessening dependence on customary input styles like keyboarding.

II. **Multifunctional Capabilities:**

The virtual assistant could potentially assist users with diverse tasks through composing correspondence, exchanging communications, and accomplishing other text-focused responsibilities.

III. **Innovative Marketing Opportunities:**

Through leveraging a virtual assistant's diverse talents, enterprises can craft promotional campaigns that are lively, interactive and truly captivate crowds - dynamically championing offerings in an engaging fashion.

IV. **Educational Enhancement:**

The virtual assistant can revolutionize education and training by providing interactive simulations and educational content.

7.Conclusion

Ultimately, the completion of our voice-based virtual personal assistant endeavor serves as a noteworthy landmark in the progression of human-machine communication through the application of natural language processing and speech identification innovations. By leveraging robust speech recognition systems and natural language understanding techniques, our assistant offers users a seamless and intuitive experience, streamlining daily tasks, improving productivity, and enhancing accessibility across various platforms.Moreover, the integration of facial recognition and OTP-based access verification enhances security, ensuring authorized user entry.While gazing ahead into the

http://doi.org/10.36893/JNAO.2024.V15I1N01.334-348

future, myriad chances for further refinement and growth await us, such as incorporating new capabilities, honing presently existing functions, and optimizing for diverse usage cases, cementing our dedication to progressing virtual assistant innovations and ensuring remarkable user involvements.

Reference

[1] Rabin Joshi, Supriyo Kar, Abenezer Wondimu Bamud and Mahesh T R, (2023). Personal A.I. Desktop Assistant, 2(2), 54-60. ISSN: 2583 5343 DOI: 10.59461/ijitra.v2i2.58 https://www.ijitra.com/index.php/ijitra/issue/archive

[2] Anusha S, N Vignesh Karthik, Sampada K S. ”Comparative Study on Voice Based Chat Bots”. 2018

[3] Pandey, A., Vashist, V., Tiwari, P., Sikka, S.,&Makkar, P. Smart Voice Based Virtual Personal Assistants with Artificial Intelligence.

[4] Swapnil Saurav, Python Programing-learn and practice (2nd Edition). Ingram short title; 2018 January 1

[6] Markowitz J. Toys That Have a Voice. Speech Technology Magazine [Internet]. 2003 March [cited 2019 April 7].

[5]Sangpal, R., Gawand, T., Vaykar, S., & Madhavi, N. (2019, July). JARVIS: An interpretation of AIML with integration of gTTS and Python. In 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT) (Vol. 1, pp. 486-489). IEEE.

[6]Mittal, Y., Toshniwal, P., Sharma, S., Singhal, D., Gupta, R., & Mittal, V. K. (2015, December). A voice controlled multifunctional smart home automation system. In 2015 Annual IEEE India Conference (INDICON) (pp. 1-6). IEEE.

[7]Pandey, A., Vashist, V., Tiwari, P., Sikka, S.,&Makkar, P. Smart Voice Based Virtual Personal Assistants with Artificial Intelligence.

[8]Subhash, S., Srivatsa, P. N., Siddesh, S., Ullas, A., & Santhosh, B. (2020, July). Artificial Intelligence-based Voice Assistant. In 2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4) (pp. 593-596). IEEE.

[9]Rahul Kumar, Garima Sarupria, VarshilPanwala, Smit Shah, Nehal Shah (2020), Power Efficient Smart Home With Voice Assistant, Ieee – 49239.

[10]Jianliang Meng, JunweiZhang,Haoquan Zhao (2012), Overview of the Speech Recognition Technology, IEEE.

[11]Dhiraj Pratap Singh, Deepika Sherawat, Sonia. “Voice Activated Desktop Assistant Using Python”. 2020

[12]Anusha S, N Vignesh Karthik, Sampada K S. ”Comparative Study on Voice Based Chat Bots”. 2018

[13]V. Këpuska and G. Bohouta, "Next-generation of virtual per-sonal assistants (Microsoft Cortana Apple Siri Amazon Alexa and Google Home)", 2018 IEEE 8th Annual Computing and Communi-cation Workshop and Conference (CCWC), pp. 99-103, 2018.