OPEN AI DESKTOP ASSISTANT

Prof. Dewendra Bharambe
Computer Department of Anantrao Pawar College of Engineering And Research, Parvati, Pune.
Savitribai Phule Pune University Pune.
dewendra.bharambe@abmspcoerpune.org

Mr. Atharav Kiran Maindarge
Computer Department of Anantrao Pawar College of Engineering And Research, Parvati, Pune.
atharavmaindarge9969@gmail.com

Mr. Omkar Vijay Naik
Computer Department of Anantrao Pawar College of Engineering And Research, Parvati, Pune.
naikomkarvijay@gmail.com

Mr. Satyam Ashok Shinde
Computer Department of Anantrao Pawar College of Engineering And Research, Parvati, Pune.
shindesatyam72@gmail.com

Abstract*—Realizing natural human-machine communication is the main objective of artificial intelligence (AI), a popular technology. To increase human-machine interaction, many IT-based businesses have also made use of conversation network technology to develop different kinds of Virtual Personal Assistants. Examples of these include Alexa, Cortana, Google Assistant, Siri, and many others. We createda virtual assistant that works similarly to Microsoft's "Cortana" voice assistant. It uses Python to carry out simple operations on the Windows platform, following instructions. Python is utilized in this case as a scripting language because of its extensive library, which is used to carry out commands. A customized virtual assistant identifies and interprets theuser's voice using Python libraries. Voice assistants are an amazing development in the field of artificial intelligence that has the potential to change people's lives in many different ways. When the voice-based assistant first appeared on cellphones, it became widely used. It was universally acknowledged. Voice assistants were formerly mostly found incomputers and smartphones, but they are currently becoming more and more common in smart speakers and different homeautomation systems. Numerous technologies appear to be becoming smarter in their unique ways, enabling them tohave basic linguistic conversations with people. Desktop voice assistants are computer programs that recognize speech patterns and respond via an integrated speech system. This essay will describe the various types of voice assistants and discuss the main drawbacks and difficulties they have. This study discusses how to construct a voice-based assistant without using cloud services, which would encourage the development of such gadgets in the future.

Keywords: Automation, Python, Speech Recognition, Voice Assistant, Smtplib.

I. INTRODUCTION

This study discusses how to construct a voice-based assistant without using cloud services, which would encourage the development of such gadgets in the future. Keywords: Speech Recognition, Voice Assistant, It makes sense to give it some input, but what if the data is your voice rather than text entered conventionally? What happens if you speak with the computer, give it instructions, and simply want it to interact with you like a personal assistant would? What happens if the system does more for the user than just show them the optimal result? What if it gives them advice on a wise choice as well? Voice commands provide an easy way to interact with the machine, making it a revolutionary method ofhuman-system interchange. We need to use an API that translatesvoice messages into text messages to comprehend the input to achieve this. Numerous businesses, such as Apple, Google, and Amazon, are working to achieve this in a universal. Numerous businesses, such as Google, Amazon, and Apple, are working to achieve this on a global scale. Isn't it wonderful that you can create reminders by just saying, "set an alarm" with "wake me up at..." or "remind me to "? Realizing how important this was, we made the decision to develop aplatform that can be put anywhere in the community and that anyone can ask for help with by just having a conversation with it. Moreover, you can use WiFi to connect two comparable devices so they can communicate with one another in the future. With frequent alerts and updates, these devices can help you perform better and can be very helpful for daily use

If not, why would we need it? Our voice is evolving into a more effective input method than a simple enter key. Users can access a wide range of applications and services from all operating systems. The most popular iPhone app is called "SIRI," which lets users interact with their phones by speaking to them or usingvoice commands. An analogous app, called "Google Assistant," was also developed by Google and is available for Android smartphones. That application does, however, rely significantly on Internet connections. The suggested system, however, can function both with and without Internet connectivity. It processes user input—such as speech or text—and returns the results in a variety of formats, including an action that needs to be taken. Voice-activated home automation systems have the potential to improve living conditions and simplify daily chores for users. For people with disabilities in particular, voice control in energy-efficient buildings offers a lifestyle that was previously unattainable. Voice-activated system implementation may have some advantages, including helping with work-related tasks. A voice assistant appears to be a computerized program agent that uses voice control technologies to carry out tasks or offer services to a person. Currently on the market a

a number of virtual voice assistants, including Google Assistant, Samsung Bixby, Apple's Siri, Microsoft's Cortana, Amazon's Alexa, and many more.

An automated program that carries out tasks or services that the user designates for it using different commands is called a voice-based assistant. Software jargon refers to the software agent that can be accessed via live chat as a "chatbot," and it falls under the "Digital agent" category. The same class of voice-based assistants is capable of comprehending and reacting to social speech.

II. IMPORTANCE OF TECHNOLOGY

The Open AI Desktop Voice Assistant is a technological marvel that revolutionizes our relationship with computers. It givesusers smooth, hands-free control over their digital worlds, increasing productivity and accessibility. This technology is an example of how artificial intelligence (AI) can revolutionize daily productivity and convenience. It also shows how technology is constantly changing and how this will affect how people interact with computers in the future.

III. LITERATURE REVIEW

On the most significant comment and speech, S. Subhash, P. N. Srivatsa, S. Siddesh, A. Ullas, and B. Santhosh have written. One significant emerging feature that is changing people's lives is voice control. Voice assistants are frequently found in laptops and smartphones. The operating systems that can identify human speech and respond with integrated voices are known as AI-based voice assistants. The audio from the microphone is captured by this voice assistant, which translates it into text before sending it via Google Text-to-Speech (GTTS). The play sound package of the Python programming language is used to play the audio file created by the GTTS engine from text [1].

Speech analysis is commonly conducted in tandem with pitched analyses, as indicated by T.-K. Kim the AI assistant system using Google Assistant, a representative service of open API Artificial Intelligence, and the conditional auto-run system, IFTTT (IF This, Then, That) was designed. It cost-effectively implemented the system using Raspberry Pi, a voice recognition module, and open software. The proposed system is expected tobe applied to various control systems based on voice recognition[2].

Humans communicate primarily through speech, according to S. M. Felix, S. Kumar, and A. Veeramuthu. The need for independent living is acknowledged in today's sophisticated, high-tech world when it comes to visually impaired individuals, whose primary issue is social restriction. They suffer in strange places with no assistance from humans. Since most tasks require visual information, those who are visually impaired are at a disadvantage because they lack access to important environmental information. Recent developments in inclusive technology allow for the expansion of the assistance provided to those who are visually impaired. The goal of this project is to use artificial intelligence, machine learning, image processing, and text to assist those who are blind or visually impaired.[3] Human life is positively impacted by the development of the Internet of the Internet of Things (IoT), as noted by H. Isyanto, A. S. Arifin, and M. Suryanegara. One example of an Internet of Things application that makes human activities easier is a smart home. Physical abnormalities are a problem because they limit the activities available to people with disabilities. To accommodate individuals with disabilities, this paper suggests a design for an Internet of Things (IoT) smart home application that includes a voice-activated remote control device. With the aid of voice commands and smart home control systems, people with disabilities can now operate their electrical appliances, including TVs, lights, and fans, without having to move to switch them on or off. Voice recognition on electrical equipment makes use of the smartphone app Google Assistant. [4]. Personal voice assistants, or PVAs, are being used more often as interfaces to digital environments, according to P. Cheng and U. Roedig. You can interact with phones, smart homes, and cars by using voice commands. The number ofsmart speakers, like Google Home and Amazon's Echo, hasincreased by 78% to 118.5 million in the United States alone, and 21% of Americans own at least one of these devices. The security and privacy of PVAs have become a top priority for users, manufacturers, and policymakers due to society's growingreliance on them. As a result, there has been a noticeable surge in research efforts in recent years addressing PVA security and privacy. Although some security and privacy studies relevant to PVAs date back to their earlier rise [5]

M. D. Babakerkhell, D. Pal X. Zhang and associates explained a Recently, there has been an increase in the use of personal voice assistants such as Google Assistant and Amazon Alexa. In order to guarantee the sustained prosperity and extensive distribution of these products, it is crucial to assess their ongoing use scenario rather than the original adoption goal. The majority of studies use an expectation-confirmation technique to assess the continuance usage scenario. However, in this work, the utilitarian and hedonistic attitudes of the users toward the ongoing usage scenario are assessed using a user engagement-based methodology. Contextual constructs like satisfaction, privacy risk, and trust are added to this. Little empirical data currently exists to support user engagement with voice assistants. [6] They defined and evaluated the semantic analyzer for "Marathi Language" during one of their projects. They began by devising a "boos trappable" encrypting approach that works even during function f, which is the technique's particular decryption function. The study found that Marathi has a high level of correctness, with consistent derivational standards when using Finite State Systems to demonstrate language in a comprehensive fashion. Because Marathi has challenging semantics, the clustering of post places and the formation of FSA are among the most important aids [6].

Speech-based intelligent personal assistants (sIPAs) have the potential to enhance the quality of life for senior citizens, but their adoption by this demographic is hindered by a number of usability issues, according to M. U. Islam and B. M. Chaudhry. In order to learn about the experiences of fourteen older adults with these devices, we conducted a semi-structured interview study. Two overarching themes emerged from the inductive and deductive coding analysis of the gathered data: the use of sIPA and concerns about sIPA. "Usage of sIPA" describes the various ways that participants intended to use their sIPAs going forward as well as how they were currently using them. The article

"Concerns regarding sIPA" describes the various usability issues that participants were having with these gadgets. Our results lead us to propose that sIPAs for adults [7].

Meftah, Alashban, Alotaibi, Y. A., and Selouani, S. A. Described A many-to-many non-parallel generative adversarial network (GAN) voice conversion (VC) model that has shown promise in style conversion tasks is the StarGANv2-VC model. The objective of this research was to examine the model's diversity and scalability for English emotional voice conversion (EVC) among various speakers and emotional states. We conducted several experiments with an Emotional Speech Database (ESD). These included a gender-independent multi-speakers-multi-emotions experiment, a single speaker-multi-emotion experiment, and a multi-speakers-multi-emotions experiment. Additionally, we evaluated the impact of training set size and contrasted the StarGANv2-VC model's performance with that of a CycleGAN model. According to our research, the StarGANv2-VC model successfully translated the voice's pitch for each of the four emotions (happy, sad, angry, and neutral). The model was more successful at converting voices for multiple speakers than at converting multiple emotions to multiple speakers. In this field, more investigation is required. Mel-frequency cepstral distortion (MCD) and root-mean-square error (RMSE) for spectrum and prosody, respectively, were used to objectively evaluate the quality of the converted speech. Additionally, we used a convolutional recurrent neural network (CRNN) to perform cross-emotion recognition.

[8] P. Burggräf et al Stated Voice assistants have manifested their existence in the vehicle over the last decade. Advancements in voice technology allow for new customer-centric scenarios. Eight dialog use cases and two interaction types were examined. The focus of this work is to answer the research question to which extent do users prefer task-oriented multi-turn dialogs over question-and-answer single-turn dialogsin certain driving situations. In 2020, a three-step online survey was conducted to gather participants' opinions on their preferences for an assistant's interaction type, use case, and perceived usefulness and pleasantness. The survey found that users favored multi-turn conversations over single-turn conversations in all defined use case scenarios. Additionally, the authors identified the changed driving situation due to factors such as electrification, automation, and connectivity as future challenges for the development of voice assistants in the automotive context.progress in autonomous driving and the focus on an integration of the voice modality as a direct function should be considered. [9]. Burggräf P et al. Declared Over the past ten years, voice assistants have become a feature of cars. New customer-centric user scenarios are made possible by ongoing technological advancements in the fields of voice recognition and user interactions. Two types of interactions and eight dialog use caseswere thoroughly analyzed in the work that followed. The goal ofthis work is to provide an answer to the research question: In what driving scenarios do users favor task-oriented multi-turn dialogs over question-and-answer single-turn dialogs? In 2020, participants completed a three-step online survey in which questions about the assistant's interaction type, use case, perceived usefulness, and pleasantness were asked. For every use case scenario that was specified, the authors discovered that users favored multi-turn talks over single-turn conversations. The altered driving environment brought about by advancementsin autonomous driving and the emphasis on integrating the voice modality as a direct function should be taken into consideration as additional challenges for the development of voice assistants in the automotive context in the future.[10]

IV. RESEARCH METHODOLOGY

- METHODOLOGY:

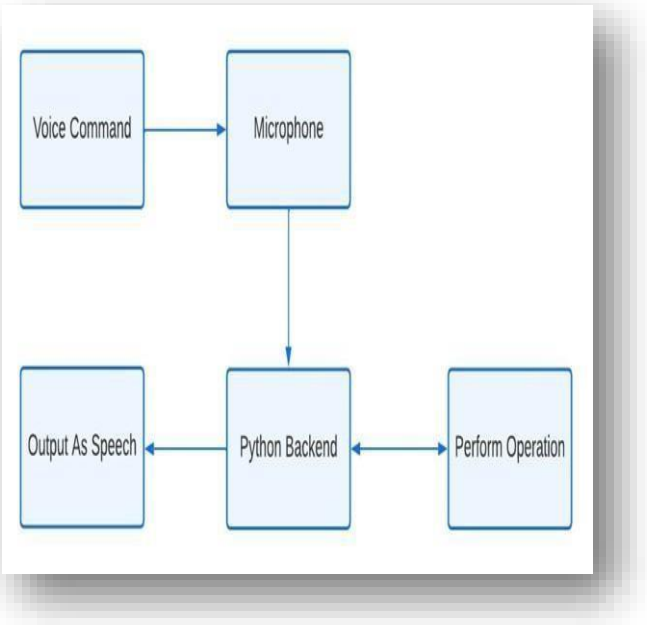

Figure 1: Basic Workflow of Model

A. Python Speech Recognition

First, the device uses a Python module speech recognition algorithm to translate user speech input into text. Using user-provided voice input, we are able to retrieve texts from specialized corpora organized on the research center's computerized network server. These texts are stored momentarily in the computer system before being sent to the speech recognition module in Python. After that, the similar text is accepted and fed into the central processor.

B. Python Backend

Python's backend parses the response from the speech recognition module to determine whether the speech or command result is a Context Extraction, Send Mail, API Call, or System Call. The information is then transmitted to the Python server's backend in order to provide the user with the pertinent outcomes.

C. API Call

Application Programming Interface is referred to as API. All that the API is a software-based interface that facilitates the establishment of a connection between two disparate systems that are located apart. In many other words, an API appears to be the

intermediary that transmits your request to the source location and then delivers the answer back to you.

D. Context Extraction

The process of extracting structured data from unstructured or semi-structured system materials is known as context extraction, or CE. Most of the time, this type of work involves interpreting texts that are readable by humans using natural language processing. TEST RESULTS for context extraction can be observed in the visual report generation activities that are currently underway, including automatic annotation from audio, images, and video and content retrieval.

E. System Calls

It seems that a software program makes a system call to request a service from the kernel of the operating system running on the computer. Examples of hardware-centric operations include the creation and execution of new processes, accessing hard drives, and facilitating communication between core kernel functions like task scheduling. System calls are used to facilitate communication between a process and the operating system.

F. Python Text-To-Speech using Pyttsx3

Text-to-speech refers to a system's ability to read aloud the supplied text (TTS). The written file is first moved to a lexical representation, then transformed into an output waveform that a TTS Engine can use to produce a sound file. TTS engines are available from third-party authors in a range of languages, dialects, and specialized vocabulary.

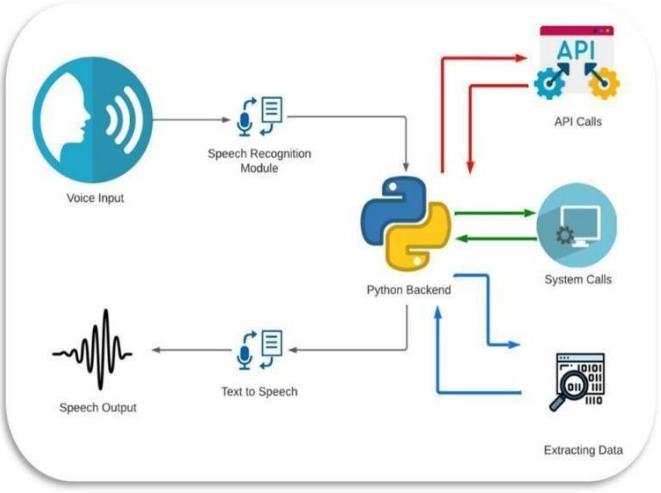

Figure 2: Detailed Workflow of Model

ALGORITHM

Several crucial steps are involved in the Open AI Desktop Voice Assistant's algorithm:

1. Voice Input: Using a microphone, the system first records user voice input.
2. Speech Recognition: To translate spoken words into text, the recorded audio is run through a speech recognition module.
3. Natural Language Processing (NLP): Next, to ascertain the user's intent and extract pertinent commands or queries, the text is examined using NLP techniques.
4. System Interaction: The assistant works with the desktop computer to accomplish tasks like launching programs, conducting web searches, and supplying information based on the user's input.
5. Voice Output: Text-to-speech (TTS) technology is used to generate responses or actions and then convert them back into speech.
6. User Feedback: The user receives the response from the assistant, and future interactions can be enhanced by taking into account the input they provide.
7. Constant Learning: Over time, the system improves its performance by continuously learning from and adapting to user preferences and speech patterns.

ADVANTAGE OF PROPOSED MODEL OVER EXISTING MODEL

Compared to current models, the Open AI Desktop Voice Assistant has many advantages. First, it uses cutting-edgemachine learning and natural language processing techniques to achieve better accuracy and comprehension of user commands. Asa result, voice interactions become more dependable and effective. Second, the system can adjust responses to specific user preferences and changing needs thanks to its flexibility and ongoing learning capabilities, which improve the user experience as a whole. Furthermore, the suggested model prioritizes user privacy and data security, giving users peace of mind at a time when worries about data protection are growing. Last but not least,the dedication to consistent upgrades and developmentsguarantees that the voice assistant stays on the cutting edge of technology, offering cutting-edge features and exceeding user expectations. In general, the suggested model offers a more intelligent, customized, and safe user experience, marking a substantial advancement in voice assistant technology.

RESULTS AND DISCUSSION

It appears that a virtual assistant is a rapid and effective helper. It's a piece of software that can interpret instructions and complete the tasks assigned to it by the client. Virtual assistants use natural language processing (NLP) to link text or speech input from users with commands that can be performed. With thehelp of a virtual assistant, you can use your instructions to operate your devices, such as laptops or PCs. It saves timebecause it is a quick process. Your virtual assistant will always be there for you and can quickly adapt to changing needs becauseit works for you and completes tasks at predetermined times. Your virtual assistant would be available to help you and others, like family and colleagues, if their workload allows.

CONCLUSION

FINAL SAY We discussed a Python voice-operated assistant in this paper. This program performs basic tasks such as opening desktop applications, playing music, checking the weather, and searching Wikipedia to assist you with daily chores. The current system's functionality is limited to the use of specific applications. Subsequent versions of this assistant will incorporate artificial intelligence, which will enhance suggestions for controlling adjacent devices via the Internet of Things, akin to what Amazon's Alexa accomplishes.

ACKNOWLEDGMENT

The Open AI Desktop Voice Assistant would like to express its gratitude to the committed group of engineers, researchers, and developers whose never-ending work made this project possible. We would especially like to thank our users for their insightful criticism and encouragement, which helped to improve the project. We express our gratitude to the open-source community for its invaluable contributions and innovations. Furthermore, we thank mentors, colleagues, and the larger AI community for their advice and support, which have been essential to the project's success.

REFERENCES

[1] S. Subhash, P. N. Srivatsa, S. Siddesh, A. Ullas and B. Santhosh, "Artificial Intelligence-based Voice Assistant," 2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4), London, UK, 2020, pp. 593-596, doi: 10.1109/WorldS450073.2020.9210344.

[2] T. -K. Kim, "Short Research on Voice Control System Based on Artificial Intelligence Assistant," 2020 International Conference on Electronics, Information, and Communication (ICEIC), Barcelona, Spain, 2020, pp. 1-2, doi: 10.1109/ICEIC49074.2020.9051160.

[3] S. M. Felix, S. Kumar and A. Veeramuthu, "A Smart Personal AI Assistant for Visually Impaired People," 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2018, pp. 1245-1250, doi: 10.1109/ICOEI.2018.8553750.

[4] H. Isyanto, A. S. Arifin and M. Suryanegara, "Design and Implementation of IoT-Based Smart Home Voice Commands for disabled people using Google Assistant," 2020 International Conference on Smart Technology and Applications (ICoSTA), Surabaya, Indonesia, 2020, pp. 1-6, doi: 10.1109/ICoSTA48221.2020.1570613925.

[5] P. Cheng and U. Roedig, "Personal Voice Assistant Security and Privacy—A Survey," in Proceedings of the IEEE, vol. 110, no. 4, pp. 476-507, April 2022, doi: 10.1109/JPROC.2022.3153167.

[6] D. Pal, M. D. Babakerkhell and X. Zhang, "Exploring the Determinants of Users' Continuance Usage Intention of Smart Voice Assistants," in IEEE Access, vol. 9, pp. 162259-162275, 2021, doi: 10.1109/ACCESS.2021.3132399.

[7] M. U. Islam and B. M. Chaudhry, "A Framework to Enhance User Experience of Older Adults With Speech-Based Intelligent Personal Assistants," in IEEE Access, vol. 11, pp. 16683-16699, 2023, doi: 10.1109/ACCESS.2022.3230151.

[8] A. H. Meftah, A. A. Alashban, Y. A. Alotaibi and S. A. Selouani, "English Emotional Voice Conversion Using StarGAN Model," in IEEE Access, vol. 11, pp. 67835-67849, 2023, doi: 10.1109/ACCESS.2023.3292003.

[9] P. Burggräf et al., "Preferences for Single-Turn vs. Multiturn Voice Dialogs in Automotive Use Cases—Results of an Interactive User Survey in Germany," in IEEE Access, vol. 10, pp. 55020-55033, 2022, doi: 10.1109/ACCESS.2022.3174592.

[10] D. Pal, M. D. Babakerkhell, B. Papasratorn and S. Funilkul, "Intelligent Attributes of Voice Assistants and User's Love for AI: A SEM-Based Study," in IEEE Access, vol. 11, pp. 60889-60903, 2023, doi: 10.1109/ACCESS.2023.3286570.