IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

VISION: THE DESKTOP VOICE ASSISTANT

[1]Pawan R. Zadgaonkar, [2]Abhishek R. Waghate, [3]Sakshi S. Nivalkar, [4]Pooja A. Kondvilkar, [5]Mrunmayee Hatiskar

[1]Bachelor of Engineering in computer, [2]Bachelor of Engineering in computer, [3]Bachelor of Engineering in computer, [5]Bachelor of Engineering in computer

[4]Assistant Professor of Mumbai University

Rajendra Mane College of Engineering &Technology Ambav (Devrukh), Ratnagiri, India

Abstract: The Desktop Voice Assistant marks a significant evolution in human-computer interaction, incorporating advanced natural language processing and voice recognition technologies into desktop computing environments. This transformative software allows users to control their computers effortlessly through spoken commands, promoting a hands-free and intuitive computing experience. This abstract delves into the key features, benefits, and potential applications of the Desktop Voice Assistant, underlining its contribution to improved accessibility, productivity, and overall user satisfaction within the desktop computing domain.

***Key Words*:** Desktop Voice Assistant, Python, Natural Language Processing, Voice Recognition, Human-Computer Interaction, Automation, User Experience**.**

I. INTRODUCTION

This In the contemporary, technology-dominated landscape, where digitalization is omnipresent, the ubiquity of smartphones underscores our constant connection to the digital realm. Now, envision a novel approach to interacting with your computer – a departure from conventional input methods like typing or clicking, replaced instead by the simplicity of spoken words. Enter our Desktop Voice Assistant project, meticulously crafted using the Python programming language. This intelligent assistant transcends the rudimentary display of date and time; it encompasses the ability to open applications, play music, and execute web searches, all orchestrated seamlessly through the medium of voice commands. Our overarching objective is not merely to facilitate computer usage but to redefine it, making the experience more accessible and enjoyable. Picture navigating through your desktop environment, not by manual actions but through the spoken word. This endeavor, affectionately named "VISION," represents more than a functional tool; it epitomizes the embodiment of a technologically adept friend right within your computer. It's not just about task completion; it's about fostering a hands-free and natural interaction with your desktop. Powered by advanced voice recognition technology, VISION aspires to transcend the realm of being merely a program. It aims to be a reliable companion, enhancing user productivity and infusing a sense of efficiency and enjoyment into computer usage. As we embark on an exploration of this project, we will delve into the synergistic relationship between Python and voice recognition technology, envisioning a future where conversing with your computer becomes as second nature as engaging in dialogue with a friend.

II. OBJECTIVE

The objective for a desktop voice assistant project using Python could be to create a versatile and efficient virtual assistant capable of performing various tasks such as:

i. Voice-controlled interactions: Responding to user queries and commands spoken aloud.
ii. Task automation: Automating repetitive tasks such as setting reminders, sending emails, or opening applications.
iii. Information retrieval: Providing answers to general knowledge questions or fetching specific information from the web.
iv. Integration with other applications: Interacting with other software and services to perform actions like fetching weather forecasts, managing calendar events, or controlling smart home devices.
v. Personalization: Adapting to user preferences and learning from past interactions to improve performance and user experience.
vi. Continuous improvement: Implementing features for ongoing updates and enhancements based on user feedback and emerging technologies.

III. RESEARCH METHODOLOGY

i. Speech Recognition

Foundational to a desktop voice assistant's functionality. Utilizes advanced algorithms to accurately transcribe spoken commands. Analyses audio input to identify speech patterns and convert them to text. Constant advancements enhance accuracy, accommodating diverse accents and languages. Crucial for overall effectiveness and user experience.

ii. Backend Processing

Takes over after transcribing spoken commands into text. Analyzes text to discern user intent and extract relevant information. Employs natural language processing (NLP) techniques for semantic understanding. Machine learning improves interpretation accuracy over time. Matches commands with predefined tasks for execution.

iii. Text-to-Speech (TTS) Conversion

Converts text responses into synthesized speech. Produces natural-sounding speech using TTS algorithms. Adjusts parameters like intonation and pitch for clarity and naturalness. Mimics human speech patterns for effective communication. Ongoing advancements enhance realism and expressiveness.

iv. Integration with Desktop Environment

Seamlessly integrates into the operating system environment. Accesses system resources, applications, and services for task execution. Utilizes APIs and SDKs to interact with various components. Interacts with email, calendar, browsers, and file systems. Ensures effective control and interaction with the desktop environment.

v. User Interaction and Feedback

Facilitates interaction through spoken commands and responses. Incorporates feedback mechanisms for acknowledgment and confirmation. Essential for improving accuracy and effectiveness over time. Analyzes user interactions and feedback for continuous improvement. Designed with user-centric principles for enhanced usability.

vi. Privacy and Security Measures

Implements robust measures for handling sensitive information. Utilizes encryption for data protection during transmission and storage. Implements user authentication mechanisms like voice recognition. Provides privacy settings for user control over data collection and usage. Conducts regular security audits to mitigate vulnerabilities and ensure user privacy.

IV. EXISTING SYSTEM

In the existing system, users need to be computer expert and provide precise instructions to find applications or files, which can be challenging. This waste time as users must type specific details using a keyboard, increasing manual effort. There's a need for a more user-friendly and efficient solution for searching and interacting with the computer.

Sure, here are the key points regarding existing systems for a desktop voice assistant with vision capabilities:

i. Lack of Pre-built Solutions: As of my last update in January 2022, there isn't a widely known pre-built system specifically designed for a desktop voice assistant with vision capabilities.

ii. Building from Existing Libraries: Developers typically build such systems using existing libraries and frameworks. OpenCV is commonly used for image processing and object detection, while SpeechRecognition handles speech-to-text conversion, and pyttsx3 is employed for text-to-speech synthesis.

iii. Integration and Customization: Developers integrate these libraries to create a cohesive system. Custom logic is implemented to process voice commands, analyze desktop images, and generate appropriate responses.

iv. Commercial Solutions:While general-purpose voice assistants like Amazon Alexa, Google Assistant, and Microsoft Cortana exist, they might not offer specialized support for desktop-specific vision tasks out-of-the-box.

v. Research and Academic Solutions: There might be specialized systems or frameworks developed within research institutions or academia to address specific requirements for desktop voice assistants with vision capabilities. These solutions are typically found through academic literature and research repositories.

V. LITERATURE REVIEW

i. Artificial Intelligence – Based Voice Assistant (Subash S., Prajwal N., Srivatsa, Siddesh S., Ullas A., Santosh B., IEEE Publication Year July 2020): The author has proposed a system that help us to suggest an idea for utilizing gTTs(google-text-to-speech) to convert spoken language into text, then storing these responses as distinct audio files, ultimately improving the interaction with voice assistants.

ii. Virtual assistant using python (Vedant Kulkarni, Vipul Wikar, Saurabh Patil, Swarupa Deshpande, IEEE Publication Year May 2020): The author has proposed a virtual assistant system that help to enable users to control devices and perform various tasks using voice commands, from opening apps to playing media and providing information like time and navigation.

iii. CHA: A Caching Framework for Home-based Voice Assistant Systems (Lanyu Xu, Arun Iyengar, Weisong Shi) IEEE Publication Year April 2020): The author proposes CHA, a caching system to improve voice assistants on edge devices by enhancing performance, understanding user commands, and reducing reliance on the cloud. CHA significantly speeds up responses for common voice commands, with efficient resource usage, such as on a Raspberry Pi.

VI. PROPOSED SYSTEM

i. Problem Statement

Design and develop a desktop assistant that can perform various tasks to assist users in their daily computer-related activities. The assistant should be able to understand voice and text commands, execute tasks efficiently, and provide relevant information to users. To develop a desktop voice assistant that enables users to interact with their computers through voice commands, reducing the need for manual typing, and enhancing overall desktop computing efficiency and user experience.

ii. Scope of Project

The project's scope includes developing a desktop assistant with features like opening Google, answering questions, search and play YouTube videos, sending messages, displaying maps, and providing date and time information. This involves voice recognition, NLP, voice activation, and seamless integration with desktop functions.

iii. Proposed System

Vision: The Desktop Assistant consist of,

The proposed system Desktop Assistant project is like building the brain and communication skills for a friendly computer helper. It will understand what you say, talk back to you, and do tasks like fetching information or setting reminders. The system will have customizable features so it fits your preferences, and it will keep your information safe. The goal is to make your computer more accessible and helpful, but it's a bit like teaching a computer to understand and talk like a human, which can be tricky. Overall, this system could make working with your computer much easier and open up new possibilities for everyone.

Proposed Systems are:

1. Accessing user voice through microphone:

Taking the input as speech patterns through microphone.

2. Voice recognition and conversion:

Audio data recognition and conversion into text.

3. Matching voice with predefined keyword:

Comparing the input with predefined commands.

4. Result:

Giving the desired output.

VII. DETAIL WORKFLOW:

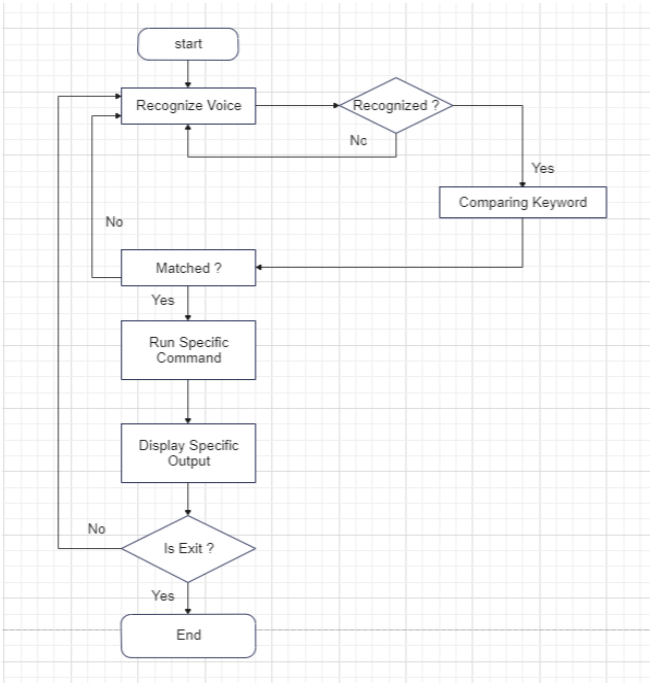


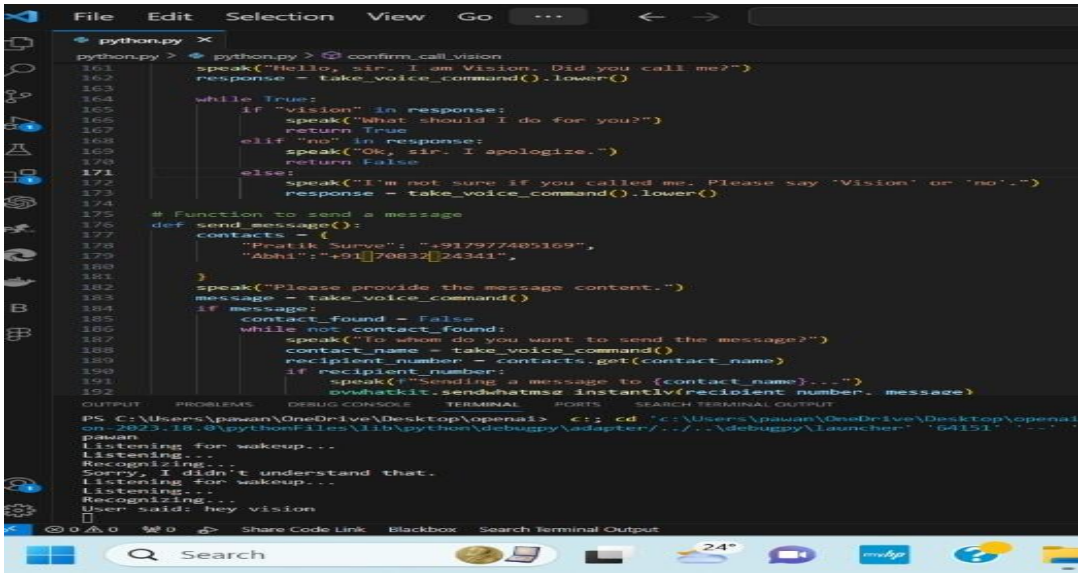
Chart -1: Flow chart

i. **Start:** This marks the beginning of the process.
ii. **Recognize Voice/Recognize?:** This is a decision point where the assistant determines whether it has successfully recognized the voice input from the user. It employs voice recognition technology to convert spoken words into text.
iii. **Comparing Keywords:** Once the voice input is recognized, the assistant compares the extracted keywords with a predefined set of keywords or commands to identify the user's intent.
iv. **Run Specific Commands:** Based on the comparison result, the assistant executes specific commands associated with the recognized keywords. These commands could involve opening applications, performing searches, sending messages, setting reminders, etc.
v. **Display Specific Output:** After executing the commands, the assistant displays the relevant output. This could include displaying search results, launching applications, showing reminders, or any other form of output corresponding to the executed commands.
vi. **Exit:** If the user requests to quit or end the interaction, Vision responds with a bye/exit message, and the process concludes.

VIII. Implementation

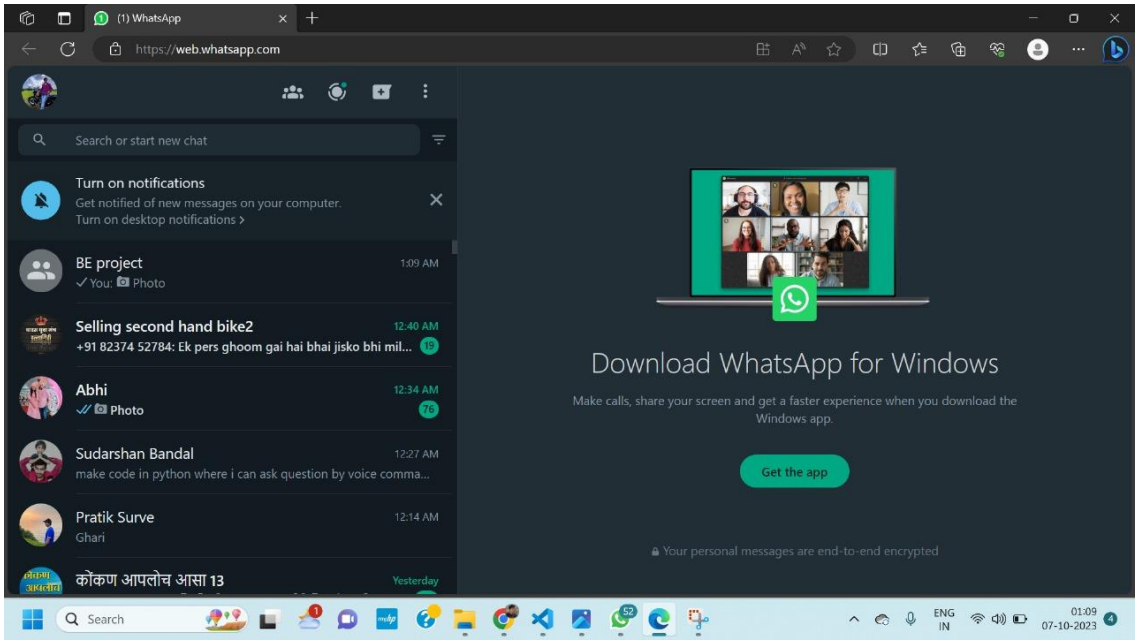
Developing a desktop voice assistant project using Python involves setting up the necessary environment by installing Python and relevant libraries such as Speech Recognition and pyttsx3. Voice recognition functionality is implemented to convert speech input into text, while user commands are processed to execute corresponding actions. Text-to-speech conversion is achieved using pyttsx3, and integration with other services such as weather updates or news headlines enhances functionality. Error handling ensures smooth operation, and optionally, a graphical user interface (GUI) can be created for a more user-friendly experience. Testing and refinement play crucial roles in ensuring the voice assistant meets user expectations and operates seamlessly.

i. Activate Vision



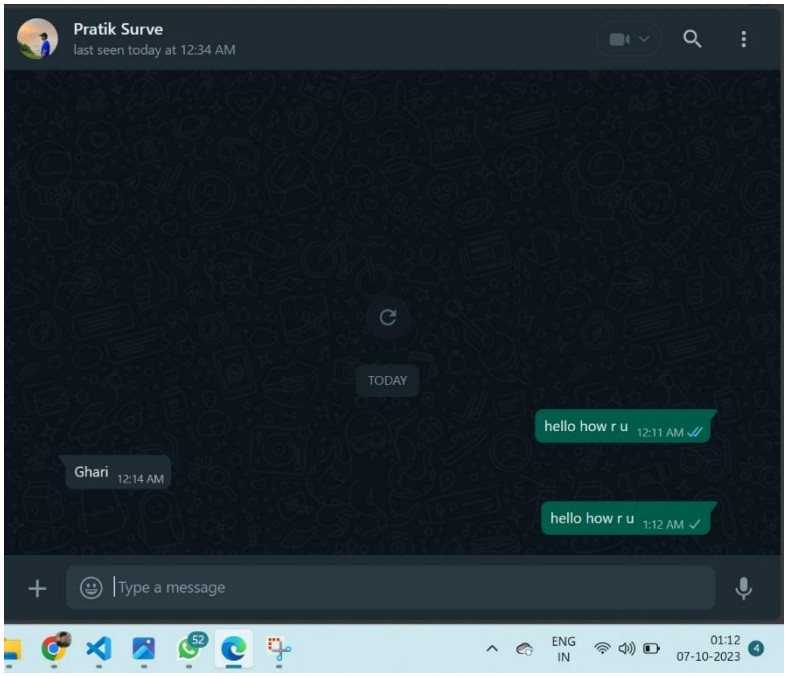
-Say the wake word "Vision" to start the assistant.
-Alternatively, click on an activation button labeled "Activate Vision."
-Vision will respond with a confirmation message once activated.

ii. Open Whatsapp



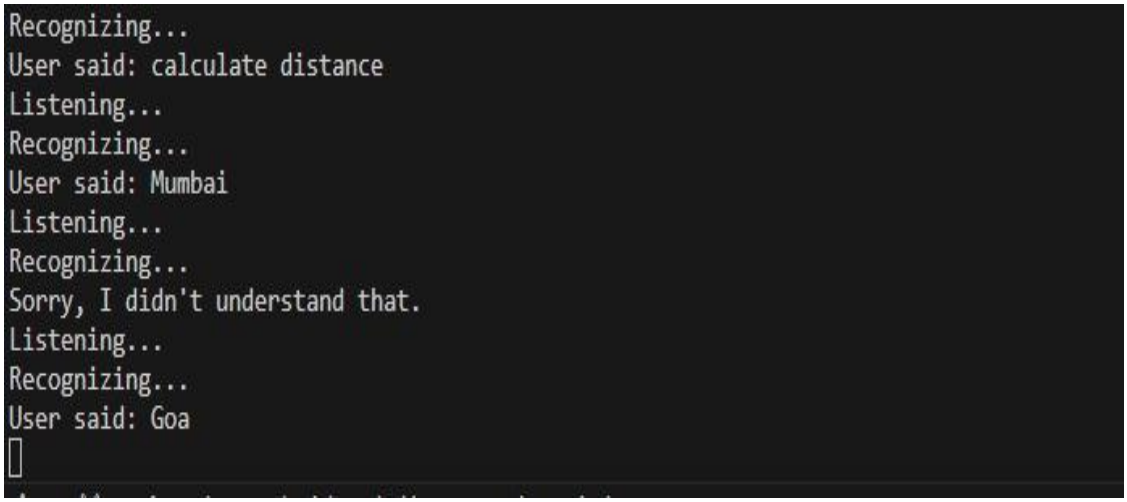
-Ask Vision to locate the WhatsApp icon on your desktop.
-Instruct Vision to simulate a mouse click to open the application.
-Once WhatsApp is opened, Vision will notify you with a confirmation message.

iii. Send Message to Whatsapp



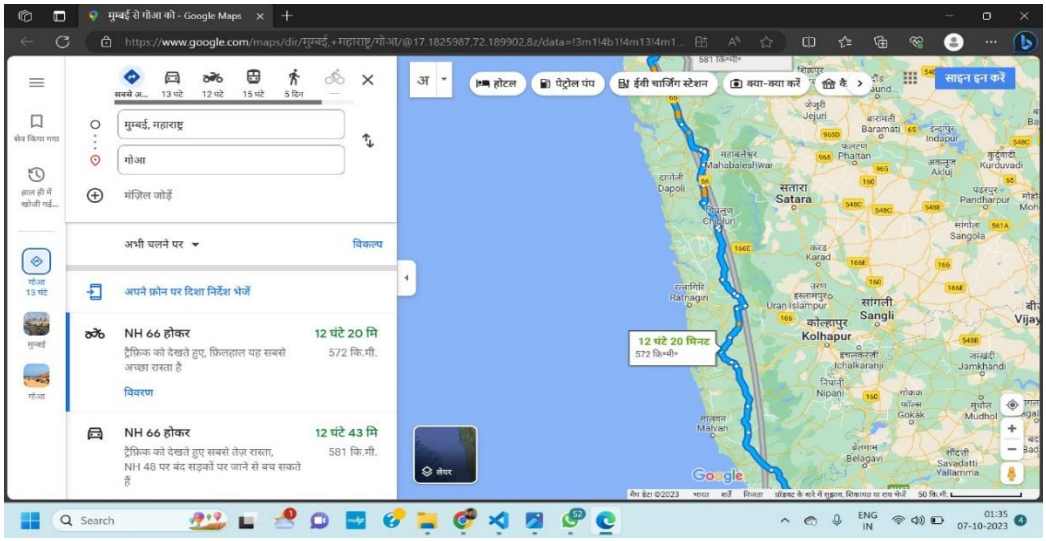
-Specify the recipient's name or contact in WhatsApp.
-Provide the exact message content you want to send.
-Vision will navigate to the chat window of the specified contact and type out the message before sending it.
-After sending the message, Vision will confirm its delivery.

iv. Calculate Distance Between Two Location



- Clearly state the names or addresses of the two locations.
-Vision will use a mapping service or API to determine the geographical coordinates of each location.
-Then, it will calculate the distance between the coordinates using a distance calculation algorithm.
-Upon completion, Vision will provide the calculated distance in miles or kilometers.

v. Show on Map



-After calculating the distance, request Vision to open a map application.
-Instruct Vision to mark the two locations on the map and display the calculated distance between them.
-Vision will then show the map with the marked locations and the distance indicated.
-Additionally, Vision can provide directions between the two locations if desired.

IX. CONCLUSIONS

Our project introduces a dynamic desktop voice assistant crafted with Python. Opening applications, exploring maps, navigating YouTube, and enabling friendly chats, this assistant showcases versatility. Its adaptive learning enhances user interactions, redefining everyday computer use. Beyond illustrating voice-enabled technology's practical application, our project signifies the evolving landscape of user-centric innovation. Looking forward, the continuous evolution promises exciting advancements, solidifying our desktop assistant's position at the forefront of seamless and inclusive computing experiences.

REFERENCES

i. Subash S, Prajwal N Srivatsa, Siddesh S, Ullas A, Santosh B., “Artificial Intelligence-Based Voice Assistant”, IEEE July 2020.

ii. Vedant Kulkarni, Vipul Wikar, Saurabh Patil, Swarupa Deshpande, “Virtual assistant using python”, IEEE May 2020.

iii. Lanyu Xu, Arun Iyengar, Weisong Shi, “CHA: A Caching Framework for Home-based Voice Assistant Systems”, IEEE April 2020.

iv. Tae-kook-Kim, “Short Research on Voice Control System Based on Artificial Intelligence Assistant”, IEEE 2020.

v. Saadman Shahid Chowdury, Atiar Talukdar, Ashik Mahmud, Tanzilur Rahman [3]Domain specific Intelligent personal assistant with bilingual voice command processing ´ IEEE 2018.

vi. Polyakov EV, Mazhanov MS, AY Voskov, LS Kachalova MV, Polyakov SV [3]Investigation and development of the intelligent voice assistant for the IOT using machine learning ´ Moscow workshop one lectronic technologies, 2018.

vii. Khawir Mahmood, Tausfer Rana, Abdur Rehman Raza [3]Singular adaptive multi role intelligent personal assistant (SAM-IPA) for human computer interaction ´ International conference on open source system and technologies, 2018.

viii. Veton Kepuska and Gamal Bohota [3]Next generation of virtual assistant (Microsoft Cortana, Apple Siri, Amazon Alexa and Google Home) ´ IEEE conference, 2018.

ix. Laura BURbach, Patrick Halbach, Nils Plettenberg, Johannes Nakyama, Matrina Ziefle, Andre Calero Valdez[3]Ok google, Hey Siri, Alexa. Acceptance relevant of virtual voice assistants ´ International communication conference, IEEE 2019.

x. Z. Zhou, X. Chen, E. Li, L. Zeng, K. Luo, and J. Zhang, "Edge intelligence: Paving the last mile of artificial intelligence with edge computing," Proceedings of the IEEE, 2019.