



International Journal of Advanced Research in Arts,  
Science, Engineering & Management (IJARASEM )

Volume 11, Issue 3, May-June 2024



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

**IMPACT FACTOR: 7.583**

# AI Based Voice Assistant

Shubh Parmar, Surajkumar Yadav

Department of MCA, Late Bhausaheb Hiray S.S. Trust's Institute of Computer Application, Mumbai, India

**ABSTRACT:** Voice assistants, leveraging advancements in artificial intelligence (AI), have become integral in various domains, from personal use in smartphones to customer service in enterprises. Voice assistant is a software that utilize artificial intelligence to take input in form of voice and then do the task accordingly. We use Various methods to convert speech into text(STT), then after processing of the text, it converts into speech(TTS). However, the smart personal assistant's study is vast. It is divided into separate branches, e.g. A computer-related environment, personal interaction with a computer as well Information systems. Various python packages is used in this project. In this paper, we have tried to present a descriptive and detailed review to provide strong support for future research. We use nlp instead of pattern identification strategies for recognizing contextual based text. It works online Python programming language is used in voice assistant. Data is stored in the app itself, it reduces the complexity of time and space. The author studies the development situation of intelligent voice assistants, and compares the differences between India and foreign voice assistants, and finally discusses the relationship between voice intelligent assistants and people's lives.

**KEYWORDS:** voice assistant, Artificial Intelligence(AI)

## I. INTRODUCTION

In today's era almost all tasks are digitalized. We have Smartphone in hands and it is nothing less than having world at your fingertips. These days we aren't even using fingers. We just speak of the task and it is done. There exist systems where we can say Text Dad, "I'll be late today." And the text is sent. That is the task of a Virtual Assistant. It also supports specialized tasks such as booking a flight, or finding cheapest book online from various ecommerce sites, and then providing an interface to book an order are helping automate search, discovery, and online order operations. Virtual Assistants are software programs that help you ease your day to day tasks, such as showing weather report, creating reminders, making shopping lists etc.

They can take commands via text or by voice. Voice based intelligent assistants need an invoking word or wake word to activate the listener, followed by the command. We have so many virtual assistants, such as Apple's Siri, Amazon's Alexa and Microsoft's Cortana. This system is designed to be used efficiently on desktops. Voice assistant software improves user productivity by managing routine tasks of the user and by providing information from online sources to the user. Voice searches have dominated over text search. Web searches conducted via mobile devices have only just overtaken those carried out using a computer and the analysts are already predicting that 50% of searches will be via voice by 2020. Virtual assistants are turning out to be smarter than ever. Allow your intelligent assistant to make email work for you. Detect intent, pick out important information, automate processes, and deliver personalized responses

### 1.1 Definition of artificial intelligence voice assistant

Artificial intelligence voice assistant is a type of speech semantic application technology which is mainly applied to speech recognition, semantic understanding, speech synthesis and other intelligent speech semantic application technology in electronic devices. Users no longer need to manually operate, and can use many operations simply by giving instructions to the voice assistant. The voice assistant understands the user's needs from the user's language instructions and completes the corresponding tasks for the user. By utilising the voice assistant, the user's experience of using electronic products such as mobile phones is better. In recent years, voice interactive technology has become the key development direction of artificial intelligence technology. Since Apple launched Siri in 2011, various electronics companies have imitated Siri's voice assistant mode and added the voice assistant function to the company's products, such as Xiaomi's Xiaoi classmate, Google Assistant, Baidu Map voice. These companies are also promoting the implementation of voice interaction technology in people's lives. According to the survey data of China Telecom Terminal Research and Test Center, the user awareness rate of voice assistant is as high as 95.6%, and the use rate is 58.6% [3]. More than half of mobile phone users use voice assistance in their daily life. Obviously, artificial voice intelligent assistant is playing an increasingly important role. In the past 60 years, the technology of voice assistants has developed rapidly. From the combination of three robust acoustic models at the beginning, the accuracy rate of voice assistants has reached the level of human stenographers, reaching 95% accuracy. Users often use the speech-to-text feature, which allows people on the go to send messages and take shorthand. Amazon launched smart speakers, which connect apps on smart phones with

furniture in daily life, opening up a broader space for voice interaction and occupying a larger market. The reason for the success of Amazon speaker is the fact that it can truly interact with dialogue, which is the most requested feature of today's social users [4]. Definitely, the understanding of natural language does not only require the voice assistant to understand the basic meaning of words, but also to understand the deep meaning of sentences through in-depth research and learning, and to connect sentences with each other

## II. ARCHITECTURE OF VOICE ASSISTANT

The architecture of a voice assistant involves several key components working in harmony to process and respond to user inputs. These components include:

### 2.1 Speech Recognition

Speech recognition is the process of converting spoken language into text. It involves acoustic modeling, language modeling, and decoding algorithms. Modern systems use deep learning techniques, particularly recurrent neural networks (RNNs) and convolutional neural networks (CNNs), to improve accuracy and robustness.

### 2.2 Natural Language Processing

NLP enables the system to understand and interpret human language. Key NLP tasks include:

- **Tokenization:** Breaking down sentences into words or phrases.
- **Part-of-Speech Tagging:** Identifying the grammatical role of each word.
- **Named Entity Recognition:** Detecting and classifying entities like names and dates.
- **Sentiment Analysis:** Understanding the sentiment behind the text.
- **Intent Recognition:** Identifying the user's intent or goal.

### 2.3 Dialogue Management

Dialogue management is responsible for maintaining the context of the conversation and managing the flow of dialogue. Techniques such as state tracking and policy learning, often implemented using reinforcement learning, are critical for effective dialogue management.

### 2.4 Response Generation

Once the intent is recognized, the system generates a suitable response. This involves text generation techniques, which can range from template-based systems to advanced generative models like GPT (Generative Pre-trained Transformer).

### 2.5 Text-to-Speech

Text-to-speech (TTS) systems convert the generated text response back into spoken language. Modern TTS systems use deep learning to produce natural and expressive speech.

## III. PROPOSED METHODOLOGY

Voice assistants take voice input and perform tasks according to the instructions provided. Working of these assistants is very simple and easy. We give some instructions to assistant in the form of audio signal then the software understands and analyse those audio signal and after that it does the tasks. The python code which we have used is also not very complex. These voice assistants can also be helpful for the peoples who cannot see. Use of python language made the execution fast and simple also. Python code has some of installer packages like speech recognition, pyttsx3, python backend, system calls.





#### IV. APPLICATIONS OF VOICE ASSISTANTS USED IN VARIOUS DOMAINS

- 1. Home Automation:** Smart Home Control- Voice assistants like Amazon Alexa, Google Assistant, and Apple Siri control smart home devices, such as lights, thermostats, security systems, and appliances.
- 2. Healthcare:** Medical Assistance-Voice assistants provide medication reminders, track patient symptoms, and offer health advice. They can also help in scheduling appointments and managing medical records.
- 3. Education:** Learning Assistance-They can help students with homework, provide explanations on various topics, and aid in learning new languages.
- 4. Automotive:** In-Car Systems- Integrated into cars to provide navigation assistance, control infotainment systems, make hands-free calls, and send messages. Safety Features- Assist in monitoring driving conditions, alerting drivers about traffic issues, and even controlling certain car functions.
- 5. Hospitality:** Guest Services-Hotels use voice assistants to provide information about amenities, order room service, make reservations, and control in-room devices.
- 6. Workplace Productivity: Meeting Management-** Schedule meetings, send reminders, and manage calendars. Voice assistants can also assist in setting up conference calls and managing tasks.
- 7. Travel and Tourism:** Travel Planning- Assist in booking flights, hotels, and rental cars. Voice assistants can also provide information on travel itineraries and weather updates.

#### V. PROBLEM DEFINITION

1. We already have multiple virtual assistants. But we hardly use it. There are number of people who have issues in voice recognition. These systems can understand English phrases but they fail to recognize in our accent. Our way of pronunciation is way distinct from theirs. Also, they are easy to use on mobile devices than desktop systems. There is need of a virtual assistant that can understand English in Indian accent and work on desktop system
2. One assistant can synthesize speech more qualitatively, another can more accurately and without additional explanations and corrections perform tasks, others are able to perform a narrower range of tasks, but most accurately and as the user wants. Therefore, there is no such assistant that can perform all the work and tasks equally.
3. As we know main problem of any voice assistant is timing, to match the timing of user what he said and give desire output on time is very difficult to perform both task simultaneously. There are some assistant like siri, cortana, alexa and etc which perform their task accurately.
4. To deal with this problem we use infinity loop which assistant will ask query every time until user say to exit assistant or say close assistant.
5. Virtual assistant should be able to model complex task dependencies and use these models to recommend optimized plans for the user. It needs to be tested for finding optimum paths when a task has multiple sub-tasks and each sub-task can have its own sub-tasks. In such a case there can be multiple solutions to paths, and the it should be able to consider user preferences, other active tasks, priorities in order to recommend a particular plan.

#### VI. DEVELOPMENT OF INTELLIGENT VOICE ASSISTANTS IN INDIA AND FOREIGN

##### 6.1 India:

##### 1. Regional Language Support

- **Multilingual Capabilities:** Voice assistants like Google Assistant, Amazon Alexa, and Apple Siri now support Hindi and other regional languages such as Tamil, Telugu, Bengali, and Marathi.
- **Local Innovations:** Indian tech companies like Haptik and Reverie Language Technologies focus on creating voice assistants tailored to India's diverse linguistic needs.

##### 2. Government Initiatives

- **Digital India:** This initiative promotes digital inclusion, driving the development of voice technologies accessible to rural populations.
- **AI Strategy:** The National AI Strategy emphasizes the role of AI, including voice assistants, in various sectors to boost technological growth.

##### 3. Industry Applications

- **E-commerce:** Platforms like Flipkart and Paytm integrate voice assistants to enhance customer experiences, facilitating tasks like search and customer support.
- **Healthcare:** Startups such as mfine and Practo use voice technologies for telemedicine, patient monitoring, and medical advice.

#### 4. Localized Innovations

- **Startup Ecosystem:** Companies like Mihup and Uniphore are developing voice AI solutions for customer service and data analytics.
- **Smart Devices:** Indian brands are launching smart devices with integrated voice assistants to meet local needs.

#### 6.2 Foreign:

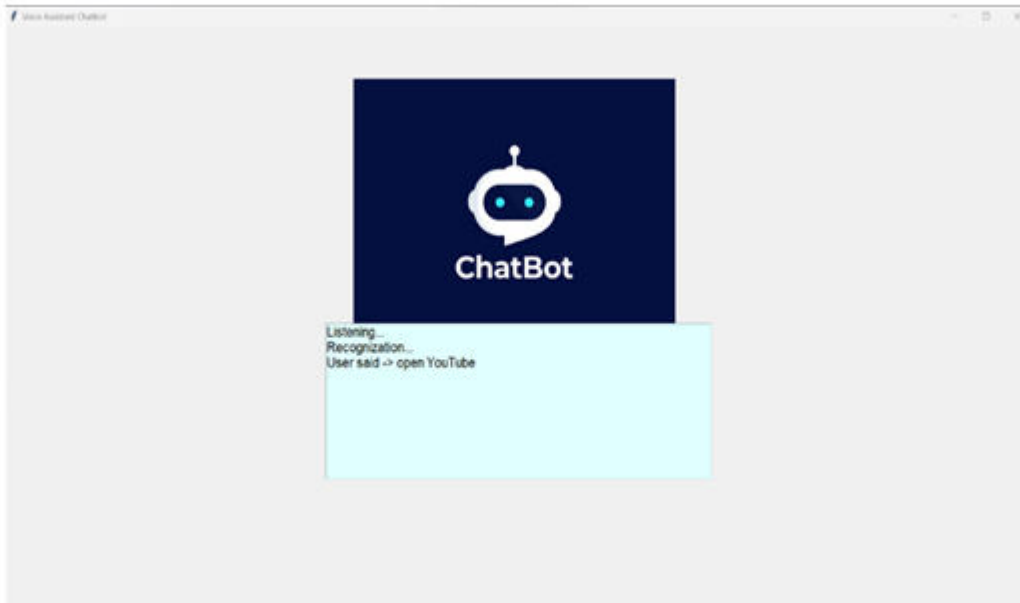
Foreign In foreign countries, four major intelligent voice assistants of Amazon, Apple, Google and Microsoft occupy the main market. Among them, Amazon's Alexa and Google's Google assistant is developing more rapidly. These voice assistants can support and accurately apply to the current mainstream demand functions, speech recognition, semantic analysis and so on. These four voice assistant companies are also constantly changing the core of their product concepts according to different user needs, striving to find suitable market groups and occupying leading positions in diverse markets[5]. As a well-known company in the smart phone market, Apple's products are primarily aimed at its own product series. Siri is mainly applied in Apple's built-in, which is conducive to Apple's promotion to users with the help of Apple's series of products. This kind of intelligent voice assistant is more suitable for users who buy Apple's companion device. Microsoft's Cortana is more associated with the PC. Cortana lets users experience when their phone and computer work together. For people who use the voice assistant for office work, Cortana will help them have a better experience at work than Siri, which is more used in daily life. Google, as the world's largest search engine, has a strong technical hand in Google assistant. Google Assistant is a virtual assistant software application designed by Google, primarily for mobile and home automation devices. Based on artificial intelligence, Google assistant can hold two-way conversations, unlike Google Now, the company's previous virtual assistant. Google assistant has the most powerful technical advantage over the previous two voice assistants, making it easier to search for information. Google assistant will analyze the user's preferences built on the user's search frequency and provide users with more valuable information. For example, a new "humming search" feature will develop by Google in 2020 will allow users to find songs simply by humming, whistling or singing. Amazon holds the largest share of the facetious voice market. In 2014, it developed Alexa, a sarcastic voice recognition system, which leads all smart voice assistant companies to become the market leader. Amazon has combined the company's appliances with a voice assistant to make a feature-rich product. Amazon no longer keeps Alexa's voice technology private, but instead provides it to third-party platforms[5]. Through cooperation with other platforms, Alexa has mastered multiple skills in a short period of time, occupying an important position in intelligent voice assistance products.

#### 6.3 Difference analysis of voice assistants in different countries

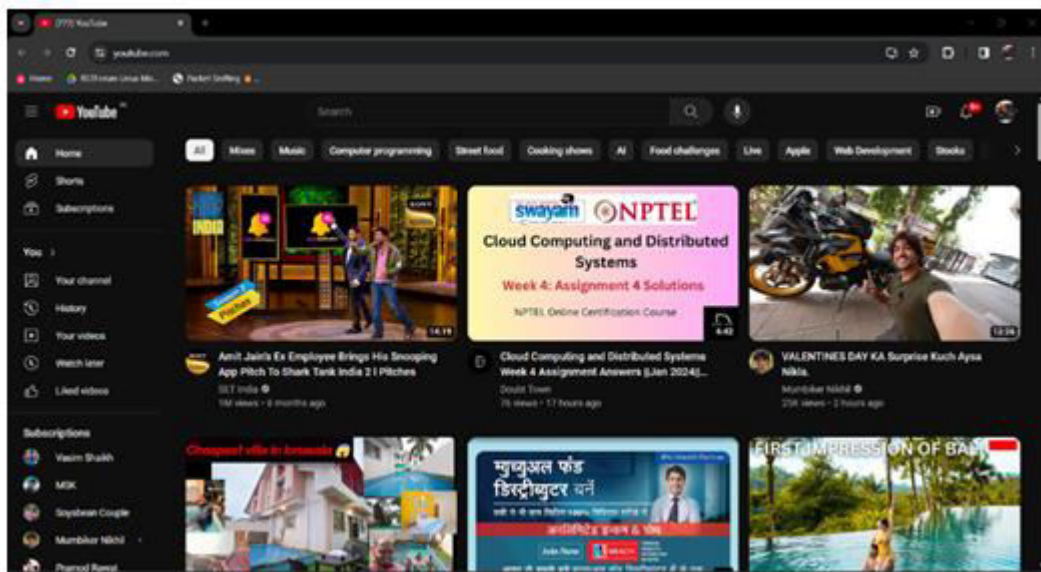
Voice assistants, such as Amazon's Alexa, Google Assistant, and Apple's Siri, exhibit distinct differences across various countries, reflecting linguistic, cultural, and regional preferences. These differences primarily manifest in language support, accents, and dialects, ensuring that voice recognition and responses are accurate and contextually appropriate for local users. For instance, voice assistants in multilingual countries like India offer support for several regional languages and dialects, whereas in Japan, assistants are optimized to understand and respond to unique Japanese honorifics and conversational nuances. Moreover, localized content integration is crucial; voice assistants in the United States might prioritize sports updates from the NFL, while those in the UK focus on Premier League football. Additionally, cultural sensitivities and social norms shape interactions; for example, the formality of responses can vary significantly between Western and Eastern countries, where politeness is deeply embedded in daily communication. Furthermore, privacy concerns and data regulations differ globally, influencing how companies deploy and manage voice assistants in compliance with local laws. These regional adaptations are essential for enhancing user experience and adoption rates, demonstrating how voice assistant technologies are tailored to meet diverse global needs effectively.

## VII. RESULT AND DISCUSSION

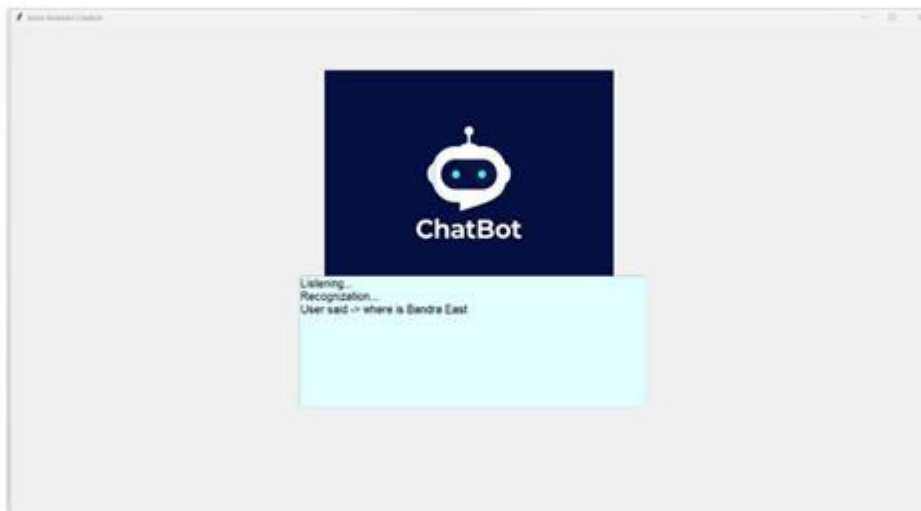
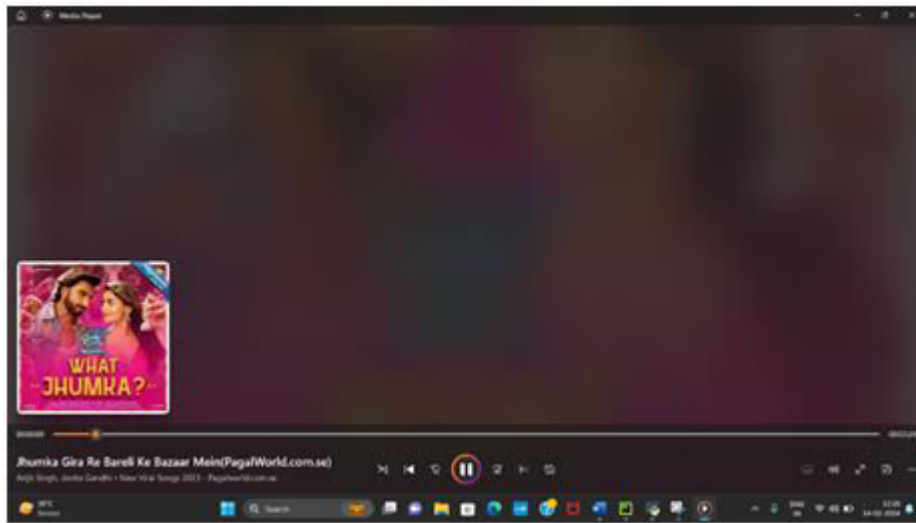
Voice assistant works fast and takes less time in execution. Voice assistant is a program that has the ability to understand instructions given by users and can do the task on the basis of those instructions. It uses system's microphone to take the verbal instructions. With voice assistant our laptops and PC's can work on our commands. It is a quick process, so saves time. Voice assistant works for you at regular intervals, therefore always accessible to you and can do the task according to the needs of user. Some of the outputs is shown in pictures given below:



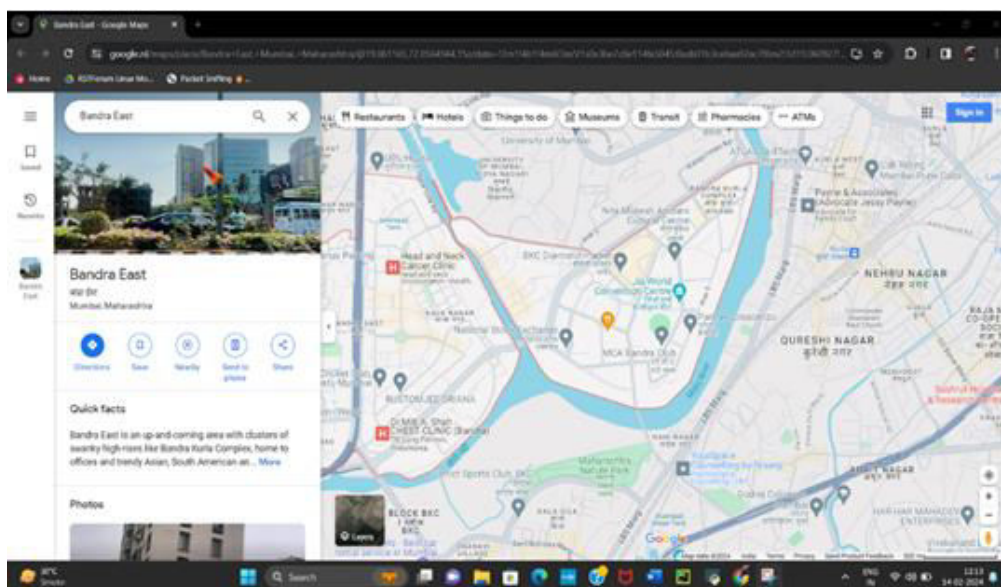
Command for opening Youtube



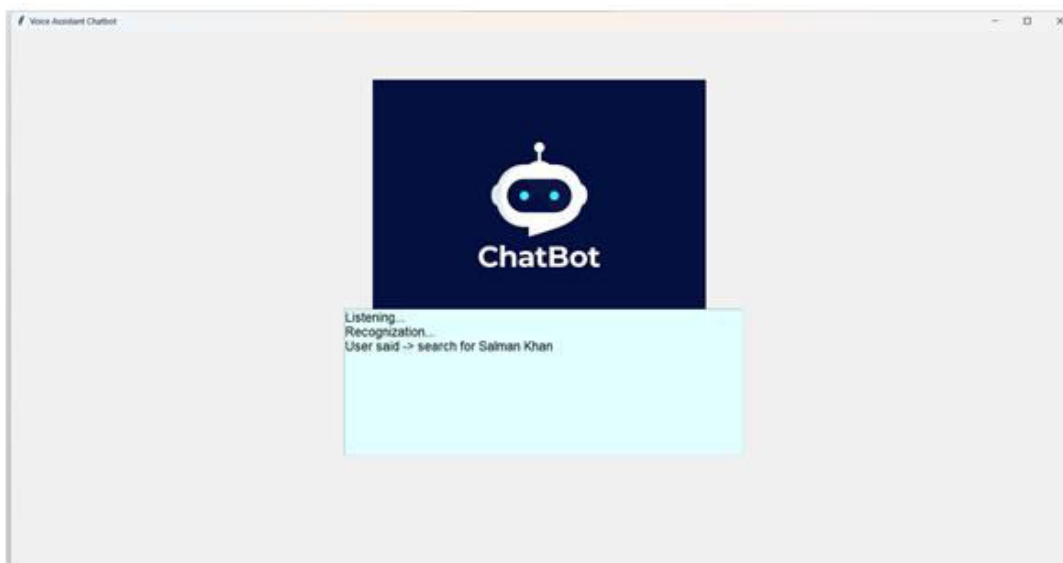
Command to Play Music



Search for Location

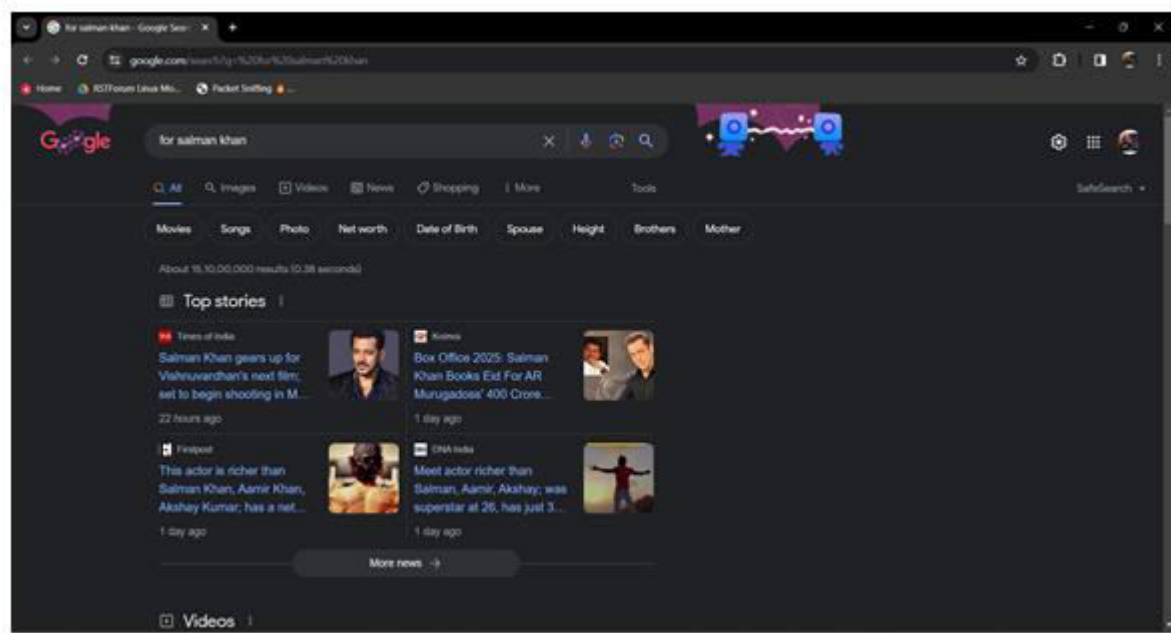


Tell us Current weather



Search anything on google





## VIII. CONCLUSION AND FUTURE WORK

We have discussed about voice assistant in the paper. Voice assistant makes life easier for people. we made a voice assistant using python programming language. It can work on various operating system like windows. Artificial intelligence is also been used for making of this project. Efficiency of voice assistants is very high due to less time consumption. Voice assistants are easy to use and they can be used at any time. As the automation process is been used in many fields now a days, voice assistant also do the automation process by reducing typing methods. Digital automation is achieved through voice assistant. It can be very useful in future as things and processes are going to be automated in future. In future voice assistant can be more efficient and accurate when we will use iot technology in these voice assistant.

## REFERENCES

1. Jurafsky, D., & Martin, J. H. (2021). *Speech and Language Processing*. Pearson.
2. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
3. Young, S., Gasic, M., Thomson, B., & Williams, J. D. (2013). POMDP-based statistical spoken dialogue systems: A review. *Proceedings of the IEEE*, 101(5), 1160-1179.
4. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv preprint arXiv:1810.04805*.
5. Ram, A., Khatri, C., Venkatesh, A., Gabriel, R., Liu, Q., Prasad, M., ... & Gabriel, R. (2018). Conversational AI: The science behind the Alexa prize. *arXiv preprint arXiv:1801.03604*.
6. Parikh, P., & Atrey, A. (2020). AI in Healthcare: Voice Assistants to Enhance Patient Outcomes. *Healthcare Informatics Research*, 26(1), 57-63.



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# International Journal of Advanced Research in Arts, Science, Engineering & Management (IJARASEM)

| Mobile No: +91-9940572462 | Whatsapp: +91-9940572462 | [ijarasem@gmail.com](mailto:ijarasem@gmail.com) |

[www.ijarasem.com](http://www.ijarasem.com)