

Implementation of Artificial Intelligence Based Sustainable Smart Voice Assistance



Mohammad Shoeb, Venkata Ratnam Kolluru, M. Naga Venkat Sai, Md. Mustafa Baig, and Shaik Razia

Abstract Artificial Intelligence (AI) technologies are being part in the human life. This article is related to the implementation of an artificial intelligence-based voice assistance system, which works by the user given commands as a request and give back the output as a response in the speech format. Core innovations are voice initiation, automated speech synthesis, speech-To-Speech, understanding common language. The proposed voice assistance system helps us to make a hand-free model, which acts as a personal assistance and mimics same as like human. Applicability and usability of the proposed model is to create an intelligent mechanism between human and computers as a natural language. Python plugins are used to train the system by using various libraries such as speech recognition, pyttsx, pyAudio. The customization of this project model makes it more flexible and freer to add new features and functionalities without disturbing the current system functionalities. It assists to eradicate the unnecessary kind of manual work required in the user life which will be performed in daily activities, not only does it operate on human instructions, but it also refers to the user based on the question or terms demanded. This Intelligent assistance communicate with the user as a result it gives a desired output as a response to the user as a voice and displays its response on the screen of the user gadget.

Keywords Artificial intelligence · Assistance · Pyaudio · RNN · Speech recognition · Text -to- speech

1 Introduction

Currently developments in computational capabilities have led to substantial evolution in AI, developments of ANN are capable of organizing human and machine natural communications. Speech, communication, gestures, facial expressions, etc.

M. Shoeb · V. R. Kolluru (✉) · M. Naga Venkat Sai · Md. Mustafa Baig
Department of ECM, Koneru Lakshmaiah Education Foundation, Guntur, AP, India

S. Razia
Department of CSE, Koneru Lakshmaiah Education Foundation, Guntur, AP, India

are gaining popularity. Among the most interested to study and popular in the world. It was the direction of communication, based on the machine’s interpretation of natural human language. A robot is taking longer time to learn from a human, but a machine learns faster to examine his behavior, acts, and to aspire to become his personal assistance. Progress has been further implementation on for a long time to build and improve these personalized assistants. There are machines that continue to grow and change, step beyond personal computers and have now established in a variety of ways. Smart computers and tablets alike. Siri, from Apple, Amazon Echo, is among the most popular voice assistants. Names of Alex, Amazon, Cortana, Artificial Intelligence In-Home Voice Assistants experienced exponential advances. However, we have so few Understanding the reasons that drive people to use those devices. Have respect to the features of the Technology, in the main hands, open, speech powered, and presentation of a voice user interface. Automated systems are growing rapidly in their way. Communicate both with a person and with oneself. New capabilities lead to the advancement of diverse mechanisms for the introduction of smart social networks into the IOT. One of the important advances in AI is the acceptance of technology as the basic language of a human being. New findings on this topic will lead to new means of natural human communication in which human and machine interaction takes place. Computers can know how to understand, alter, and express natural speech. One such tool is the voice assistant, which can be implemented with a number of other intelligent programs. Guidelines for the operation of voice assistants are set out in this article. Key limitations and weaknesses are identified. Method for creating a local voice assistance without using cloud technology that would make it possible to massively expand the applicability of those products in the future.

This paper work is carried forward with different measures by explaining. In Sect. 2 the paper reference will be discussed and the methods and technologies used in the earlier publication and the work is done by referring the various methods. In Sect. 3 the procedure to work with this model is explained and the elements used as voice inputs, voice output, speech recognition, python backend work which includes the user commands and user given quires used in the part of coding and text-to-speech conversion module. In Sect. 4 the assistance working and the interaction of the smart voice assistance is shown in the form of results to the user-end and the final assistance working performance is described in Fig. 1 there is the user

Fig. 1 Architecture for voice assistance

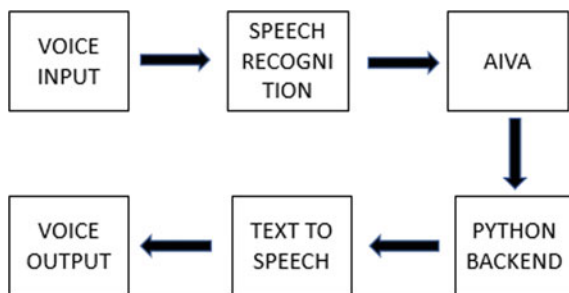


Fig. 2 Commands recognizing from user side

```
listening...  
recognizing...  
user said : open YouTube  
listening...  
recognizing...  
user said : coronavirus symptoms  
listening...  
recognizing...
```

interface commands and recognized quires as commands by the user side in Fig. 2 final display of the communicated requests and responses between human and smart voice assistance.

2 Literature Survey

McLean and Osei-Frimpong [1] Understanding the reasons that drive people to use those devices. Have respect to the particular features of the technology, in the main hands, open, speech powered, and presentation of a voice user interface, Current technological implementation models are not adequately detailed to clarify the adoption of this emerging technology. The advancement of technology and its deployment in a number of fields are not adequate to ensure market use and the exploration of the possible benefits it offers [2]. For this reason, advanced understanding of the performance factors relevant to AI-based smart goods is required from the planning stage. Due to the ability of AI technology to transform culture, AI-based smart products would have a significant effect on life. Features associated with the introduction of mobile smart technologies the mobile technology studied included Intelligent Glass and Smart Device the Innovation Adoption Paradigm was tested in a wide variety of scenarios in the field of technology acceptance research and gained substantial empiric support [3]. Model framework has also been developed, featuring a dynamic graphic assistant capable of displaying speech patterns and allowing recognition technology and recognition, facial detection and face detection for user identity [4]. Multimodal Dialog systems that process two or more combined user inputs Aspects such as voice, picture, film, touch, physical motion, head and body movement to develop the Next Generation VPAs model [5]. In order to make a new generation of robotic personal assistants as multi-domain multi-application decentralized speaker recognition [6]. This is the first addition is the Assistance Design, which consists of separate third-party programs managed by the Controller. In this view, frameworks are electronic devices that respond. Voice assistants are useful in a variety of areas, such as Education, everyday life, home appliances, etc. Voice Assistant is also helpful for those who are illiterate [7]. Get some details only by asking the secretary. Open to users, thanks to AI-based voice assistants. Voice assistants are

useful in a variety of areas, such as Education, everyday life, home appliances, etc. Voice Assistant is also helpful for those who are illiterate. Get some details only by asking the secretary. Open to users, thanks to AI-based voice assistants. Voice Assistant is developing more and more in everyday life. Most voice assistant firms are seeking to improve Connection and more about the next level functions and much of the youth began using a voice assistant in their everyday lives and from there. The product of multiple outlets shows very positive reviews. A virtual, integrated voice assistant in customized assistant creation consisting of gTTS, AIML Artificial Intelligence Markup Language, and Python-based state-of-the-art technology. It integrates the influence of AIML in the gTTS libraries and with the industry-leading Google text-to-speech conversion tool and the voice of the Male Pitch. This is a unique consequence of the inflated contribution of many contributors, such as the feasible use of AIML and its complex merger with platforms.

A virtual, integrated voice assistant in customized assistant creation consisting of gTTS, AIML Artificial Intelligence Markup Language, and Python-based state-of-the-art technology [8]. It integrates the influence of AIML in the gTTS libraries and with the industry-leading Google text-to-speech conversion tool and the voice of the Male Pitch. Iannizzotto et al. [9] design a method to test and compare proprietary speech recognition systems with open-source speech recognition systems such as Sphinx-4, such as the Microsoft Speech API and Google Speech API. In comparison, human-chatbot contact lacked a great deal of the richness of words used in people's discussions and displayed greater profanity [10]. These findings indicate that while human language skills are easily translated to human-chatbot interaction the substance and nature of such discussions are markedly different. Innovations in smart assistants and intelligent assistants [11]. Lately, home automation has attracted attention and excitement. Consumers and scientists. Virtual Assistants Allowed Speech often referred to as smart speakers provide a broad range of network-oriented services and can connect to smart services in some situations. Cloud-based applications are highly dependent on smartphones, thereby transmitting potentially confidential data to servers that are distant. Exponential growth of artificial intelligence and mobile computing offers the blind a more comfortable existence and people with vision disability [12]. This essay introduces a prototype for of a specially built voice assistant for them. Koppula and Negi [13] building a voice controlled personal assistant here. The commands of the human voice are issued to the robotic by using a smart cell phone, assistant remotely. The Robot various gestures, twists, start or stop procedures may be carried out relocating an object from one position to another. The Voice commands are processed using an online cloud in real-time. Artificial intelligence is one of the priorities of realization of normal human and computer dialogue [14]. Dialog applications, sometimes referred to as immersive systems, have been used in recent years. The fastest developing field of AI is conversational systems.

Physical connectivity with the intention of having commands or access to a computer system are now available [15]. Systems stimulated by voice or speech are a part of the culture of digital smartphones. The Automatic speech acknowledgement is an important application of artificial intelligence Technologies for artificial intelligence are starting to this is encouraged by the appearance of being consciously

used in human life, the Internet of Things, and its wide distribution IoT [16]. To connect, autonomous devices are getting smarter. With an individual as well as with themselves. New capacities contribute to development similar mechanisms for incorporating smart things into social systems the Internet of Things networks. One of the pertinent developments in Artificial intelligence is the science of natural recognition. It enhances a normal contact between humans and robots, in which the computer must learn how to comprehend the language of human modifying and engaging inside it. Speech assistants are one of those methods are used to implement many other smart systems [17]. One of mankind's biggest challenges is vision disability. To execute everyday activities, some individuals require some support. This describes a system that helps users to read messages about the world, words, letters used in postal letters, daily newspapers, and so on to cope with social life [18]. Earlier, these persons needed paper type Braille to read messages or require assistance. For greater and smoother contact with culture, environmental messages are translated to voice or audio.

Understanding the reasons that drive people to use those devices. Have respect to the particular features of the Technology, in the main hands, open, speech powered, and presentation of a voice user interface, Current technological implementation models are not adequately detailed to clarify the adoption of this emerging technology.

3 Working Procedure

A continuous network be a style of artificial neural network which can be used in speech recognition and language processes, the recurrent neural network is used to find the successive features of the data and it also used to forecast the subsequent apparent conditions, this RNN unit which is also used in deep learning and in the creation of models that mimic the behavior neurons inside the human brain. They are especially effective in the use of cases where the meaning is present. In other ways recurrent neural networks are the state of sequential data algorithms has been used by the companies like apple, google etc. you know how the google is autocomplete the feature of the reminders of the google typing or about google assistance "McLean et al. [1]". we will give our input as an audio however the audio format is sensed and recognize by the speech recognition then the pyAudio is used to convert the received audio format input into the text format to make it as the user given input to generate the output. In these a large amount of data can be observed by the machine and it helps to find the frequently used words.

While the user when using the Speech Recognition, the machine utilizes Google's online speech recognition system to translate speech data to text. Speech input given by the Users can obtain text from a microphone from a special corporation arranged. The corresponding text would then be sent and It's fed to the user as an output. In the Context Extraction the process is automatically extracting structured information from unstructured and semi-structured content. Machine-readable papers, please. In most cases, this practice involves the retrieval of human language texts by natural

means. Production of languages. Latest activities in multimedia text preparation such as automated annotation and content extraction Out of the images/audio/video may be used as test Outcome context extraction. We use Text-To-Speech refers to computers ability to translate text aloud. A TTS Engine transforms written text to phonemic text. Representation, then it transforms the phonemic representation to waveforms which can be output as an audio. TTS engines of different specifications Languages, dialects and advanced terminology are accessible from third-party publishers. In this we are also having the AI Virtual Artist it can be used as request and response medium weight is based on its uses to your response of them to get the desired input and output required by the users. Generally, Voice input and Output Speech or voice input or output is the ability of a computer to accept and translate dictation or to recognize and execute spoken commands which was given by the user. Voice recognition has gained popularity and benefit from the advent of AI and intelligent assistants such as Amazon's Alexa, Apple's Siri and Microsoft's Cortana. Speech recognition technologies allow users to communicate with devices simply by talking to it, allowing hands-free questions, alerts and other basic tasks. Import speech recognition as sr-Speech recognition is used in the python backend to detect our voice and convert it as a text by using the Pyaudio library and it is useful for taking input by voice audio control speech innovation mission which is used for speech to text and text to speech conversion and it is able to take comments by using this inbuilt microphone in other devices also helps the user through microphone taking the voice and convert it as a text it gives and response by assistance. An another library Import pyttsx3-Pyttsx3 is used by the user for doing the functionality text-to-speech translation library in Python. Unlike alternative libraries, it operates offline and is compliant with both Python 2 and Python 3. The program invokes the `pyttsx3.init()` factory function to obtain a connection to `pyttsx3`.

From Fig. 2 how the commands are taken from the input source and the response given to the output source is clearly explained and each module functioning. It is a very easy-to-use method to translate the text you entered speech. Import date-time—Date and datetime are objects in Python, because when you modify them, you are simply modifying objects, not strings, or timestamps. Whenever you control dates or hours, you need to import the datetime feature. The datetime classes in Python are assembled into five major classes. Import web browser-In Python, the web explorer module offers a high-level GUI that helps users to access Web-based documents. The web browser module will be used to open a web browser in a platform-independent manner. Comes with Python and opens a window for a particular website. Requests, please. Downloads files and websites from the Internet. Import Wikipedia—Wikipedia could be a Python library that produces it simple to look at and search Wikipedia data. Wikipedia packages the Media Wiki API in such some way that you just will focus on victimization Wikipedia data, not obtaining it. We can use the Python Wikipedia API to retrieve data from Wikipedia. To decision the strategies of the Python Wikipedia module, we want to import them victimization the subsequent instruction. Installing pyAudio for speech recognition processes PyAudio is additionally obtainable if you decide on to use audio feedback from microphones. If it isn't mounted, the library can still operate, however

the electro-acoustic transducer is going to be vague. Official PyAudio builds tend to be broken on Screens. As a result, you will notice the unofficial PyAudio builds for Windows that truly add the installer folder. Run associate installer that matches your Python version to put in PyAudio. In Debian-based distributions like Ubuntu, you'll sometimes. Install PyAudio by running `sudo apt-get upgrading python-audio python3-pyaudio`, which is able to install it for each Python two and Python three.

4 Results and Discussions

Encountering with many A.I based devices with our surroundings and environment there being an improvement in smartness and AI process. Many devices are working based on AI and machine learning, voice assistance having a built-in python environment functioning and procedure by using the existing libraries. In future we can improvise our assistance functionalities and quires as by user requirement. This can be changed or upgraded by the developer side by using the required libraries and other coding techniques it is simple and easy to modify the assistance as own choice. So, by using RNN techniques to build the assistance in which usually helps in speech recognition to communicate with devices it would give a request and response to the humans so that we can have one request and response communication between human and a device. By these technologies most of the time will be saved by using these types of gadgets you can do much easier things by virtual assistance. Many assistances in this upgrading world like Alexa, Siri and such devices which are helpful in our daily life. References and sources that are mostly used to collect information and helps in building an assistance moreover this type of devices and gadgets will be used in future as our daily life routine. Forever it will be used in the future for controlling for the home appliances, electronic devices and motor vehicles. Artificial intelligence makes a major change in our future, so that the results of sustainable voice assistance would help us in using of A.I in our daily life for the future.

From Fig. 3 we can see how the voice assistance takes the command from the user it will listen speech from the user side and convert the audio into text format for understanding and it will display the recognized speech on the screen.

5 Conclusions

Modular in design this project makes it more versatile and simpler to incorporate new functions without disrupting the existing framework functionality. Not only does it operate on human instructions, but it also refers to the user based on their quires or terms demanded. Speech by the end user, such as opening tasks and procedures. The way the user is more relaxed and feels free is to greet the user. Interacting with the smart voice assistant. The program should also remove any excessive manual work done by the customer. The whole system operates on the verbal feedback rather than

```

listening...
recognizing...
user said : open YouTube
listening...
recognizing...
user said : coronavirus symptoms
listening...
recognizing...
user said : coronavirus precautions
listening...
recognizing...
user said : heart attack symptoms
listening...
recognizing...
user said : song viral fever
listening...
recognizing...
user said : various diseases in winter
listening...
recognizing...

listening...
recognizing...
user said : Billa movie
listening...
recognizing...

listening...

```

Fig. 3 Response from the assistance in text format

the text one. And we are having a feature in this model that we can add any features or functionalities in the code we can easily adjust serviceability as user defined with the help of simple code, and there is no of any other devices additional required devices to work. Operations done by this smart system is too easy to handle and user can make this model as a feature updating device because after adding some more functionalities in future this model will work effectively in all applications.

References

1. Guitton M (2019) *Comput Human Behav* 99(5): 28–37
2. Sohn K, Kwon O (2019) Technology acceptance theories and factors influencing artificial intelligence-based intelligent products. *Telemat Inform* 101324
3. Adapa A, Nah FF-H, Hall RH, Siau K, Smith SN (2017) Factors influencing the adoption of smart wearable devices. *Int J Human-Comput Interact* 34(5):399–409
4. Iannizzotto G, Bello LL, Nucita A, Grasso GM (2018) A vision and speech enabled, customizable, virtual assistant for smart environments. In: 2018 11th international conference on human system interaction. IEEE. <https://doi.org/10.1109/HSI.2018.8431232>
5. Kepuska, V, Bohouta G (2018) Next-generation of virtual personal assistants (Microsoft Cortana, Apple Siri, Amazon Alexa and Google Home). In: 2018 8th annual computing and communication workshop and conference (CCWC). IEEE. <https://doi.org/10.1109/ccwc.2018.8301638>
6. Ekeinhox-Komi T, Bouraoui J-L, Laroche R, Lefevre F (2016) Towards a virtual personal assistant based on a user-defined portfolio of multi-domain vocal applications. IEEE. <https://doi.org/10.1109/SLT.2016.7846252>
7. Subhash S, Srivatsa PN, Siddesh S, Ullas A, Santhosh B (2020) Artificial intelligence-based voice assistant. In: 2020 Fourth world conference on smart trends in systems, security and sustainability (WorldS4). IEEE
8. Hill J, Ford WR, Farreras IG (2015) Real conversations with artificial intelligence: a comparison between human-human online conversations and human-chatbot conversations. *Comput Human Behav* 49:245–250

9. Iannizzotto G, Bello LL, Nucita A, Grasso GM (2018) A vision and speech enabled, customizable, virtual assistant for smart environments. In: 2018 11th international conference on human system interaction (HSI). IEEE
10. Chen R, Tian Z, Liu H, Zhao F, Zhang S, Liu H (2018) Construction of a voice driven life assistant system for visually impaired people. In: 2018 International conference on artificial intelligence and Big Data (ICAIBD). IEEE
11. Merugu S, Tiwari A, Sharma SK (2021) Spatial spectral image classification with edge preserving method. *J Indian Soc Remote Sens* 49(3):703–711
12. Merugu S, Reddy MCS, Goyal E, Piplani L (2019) Text message classification using supervised machine learning algorithms. *Lecture notes in electrical engineering*, vol 500, pp 141–150
13. Koppula VK, Negi A (2011) Fringe map based text line segmentation of printed Telugu document images. In: *Proceedings of the international conference on document analysis and recognition, ICDAR*
14. Koppula VK, Atul N, Garain U (2009) Robust text line, word and character extraction from telugu document image. In: 2009 2nd international conference on emerging trends in engineering and technology, ICETET 2009
15. Narayana VA, Premchand P, Govardhan A (2009) A novel and efficient approach for near duplicate page detection in web crawling. In: 2009 IEEE International advance computing conference, IACC 2009
16. Mishra A, Makula P, Kumar A, Karan K, Mittal VK (2015) A voice-controlled personal assistant robot. In: 2015 International conference on industrial instrumentation and control (ICIC). IEEE
17. Kepuska V, Bohouta G (2018) Next generation of virtual personal assistants (Microsoft Cortana, Apple Siri, Amazon Alexa and Google Home). In: 2018 IEEE 8th annual computing and communication workshop and conference (CCWC). IEEE
18. Rawat S, Gupta P, Kumar P (2014) Digital life assistant using automated speech recognition. In: 2014 Innovative applications of computational intelligence on power, energy and controls with their impact on humanity (CIPECH). IEEE