Experiment 1

Aim: Study various applications of Artificial Intelligence.

Theory :

Topic : Voice Assistance using AI

Literature Review:

(1)

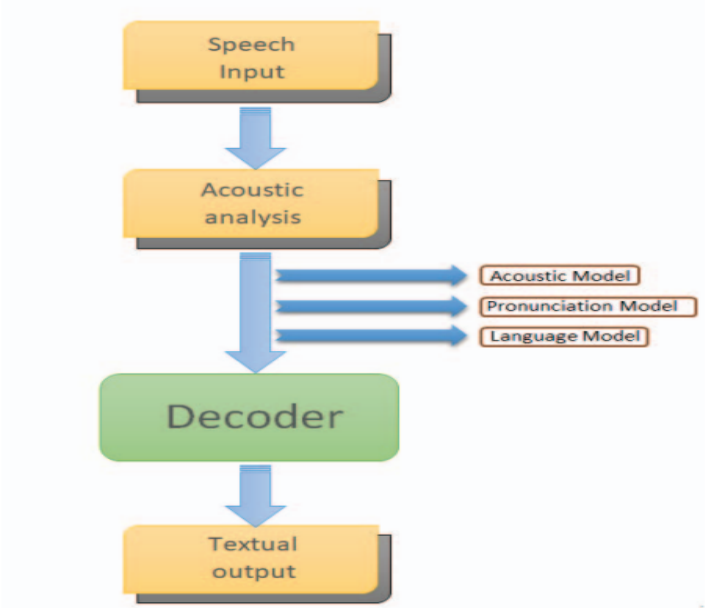
Artificial Intelligence based Voice Assistant.

Author - Subhash S, Prajwal Srivasta, Siddesh S, Ullas A, Santhosh B

Year - 2020

Abstract - Voice control is a major growing feature that change the way people can live. The voice assistant is commonly being used in smartphones and laptops. AI-based Voice assistants are the operating systems that can recognize human voice and respond via integrated voices. This voice assistant will gather the audio from the microphone and then convert that into text, later it is sent through GTTS (Google text to speech). GTTS engine will convert text into audio file in English language, then that audio is played using play sound package of python programming Language.

Introduction and Methodology - The Article explains the importance of Voice Assistance in day to day tasks and improvements to be performed.It also has differentiated between voice assistance and virtual assistance. They used python programming language to build the AI based voice assistant. They also used ASR (Automatic Speech Recognition) and gTTS engine to convert, search and process the command.



'gTTS' is basically used to convert the audio string into text. This audio string is nothing but the response which the voice assistant is supposed to give the user. The language of the text is chosen to be English, the code for English is 'en'. We save this entire function into 'tts'. We are saving this text, that is the audio file with the '.mp3' extension. Each audio file is given a random number from 1 to 20000000. The random number can be generated using the command 'random.randint()'. The whole 'mp3' file is saved under the name 'audio file'. Finally to save the file, we used the command tts.save(audio file).

Automatic Speech Recognition which is termed as ASR is the main principle behind the working of AI-based Voice Assistant.ASR systems, at first it records the speech, then the wav file has been created by the device which consists of the words it hears, later the wav file will be cleaned so that the background noise would get deleted and the volume will be normalized, then it will break down into elements and it will be analysed in sequences, then the ASR software examines these sequences and it implements statistical probability to find out the entire words and then it will get processed into text content. The better method to recognise elements is Element Recognition as it provides better results than the method of word decoding. It does not matter what kind of speech recognition software we may use, because all the work happens in its ASR. During a nutshell, at first the method starts with the device gathering audio with the source, where source is

microphone, then the Recorded speech waveforms will be sent to acoustic analysis, which will be performed on three different levels, In Acoustic Analysis, it represents that the elements were pronounced or not and what are the words which can complete these elements.That analyses the way, where how these elements are pronounced, it will check whether there is any accent or other peculiarities.Language Modelling often aimed toward finding contextual probabilities counting on what elements were captured. All the data which were recorded get processed by Artificial Intelligence without any human interaction, then the speech waveforms data is transmitted to the decoder, where it finally transforms into text for further use like command.

Observations and Findings - It has been observed that the AI-based voice assistant has been useful in many cases. It can behave like a typical search engine when needed to and can also behave like a bot. It can remember the user's name till current session and can also perform tasks under permissions. It can play or download a song, take screenshots when necessary. It can also highlight important locations specified by the user on google maps saving time and increasing the efficiency. It can give live news around the world and provide the recent and most famous results. It can be used to tell whether a password has been hacked or not. If a person is in danger, it can be used to send the user's location to the police or close relatives by giving the command “I’m in danger”.

Conclusion - Building a hands-free application which is simple yet effective is very useful for businesses. For example if a person is wearing gloves or suits as a precautionary measure in labs or any situation where it is difficult to type, they can use voice assistance to get information so that their work becomes easy. It can be used as a translator for tour guides. AI is evolving rapidly and compared to the last 2 years, AI systems have been developed rapidly more and more as it helps improve efficiency.

(2)

AI-Based Voice Assistant Systems: Evaluating from the Interaction and Trust Perspectives.

Author - Farzaneh Nasirian, Mohsen Ahmadian

Year - 2018

Abstract - Artificial Intelligence (AI) technologies are one of the new technologies with new complicated features, that are emerging in a fast pace. Although these technologies seem to be extensively adopted, people do not intend to use them in some cases. Technology adoption has been studied for many years, and there are many general models in the literature describing it. However, having more customized models for emerging technologies upon their features seems necessary. In this study, we developed a conceptual model involving a new system quality construct, i.e., interaction quality, which we believe can better describe adoption of AI-based technologies. In order to check our model, we used a voice assistant system (VAS) technology as an example of this technology, and tested a theory-based model using a data set achieved from a field survey. Our results confirm that interaction quality significantly affects individual’s trust and leads to adoption of this technology.

Introduction and Methodology - Artificial intelligence (AI) is the demonstration of intelligence by machines, which has numerous applications in various industries. Web search engines have been developed to search for information on the internet, but their limitations have led to the development of AI-based voice assistant systems (VASs) that provide intelligent search techniques. This study aims to propose an adoption model for AI-based VASs, using social exchange theory to explore the potential quality factors affecting adoption and how these factors can help VASs to be adopted. The

study is explorative and could help businesses better understand which aspects of new AI-based technologies are more important for users, leading to more effective service design and introduction.

This article discusses the development of technology and its adoption by users, with a focus on AI-based Voice Assistant Systems (VASs). The authors explore various models and theories, such as the Technology Acceptance Model (TAM), Information Systems Success Model (ISSM), Social Exchange Theory (SET), and the Diffusion of Innovation model, to better understand the factors that influence user adoption and acceptance of new technologies. The authors propose a new model that combines ISSM, SET, and HCI, with personal innovativeness as a moderator, to better describe the adoption of emerging technologies such as VASs.

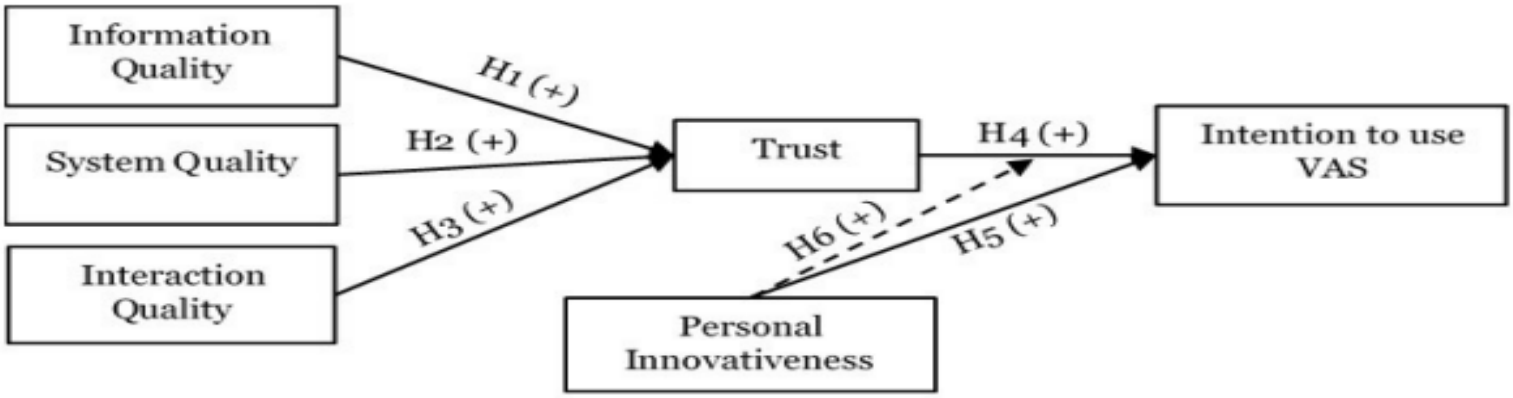


Figure 1. Research Model

Observations and Findings - The study aims to understand user preferences and their effect on the adoption of AI-based technologies. The study found that interaction quality is important for adoption of AI-based systems, such as voice assistant systems. The study also found a positive relationship between trust and personal innovativeness with intention to use. However, the relationship between information quality and system quality with trust is not supported in this study. The study provides valuable insights for researchers and practitioners, including the need to consider interaction quality in the design and development of AI-based systems, and the importance of understanding users' assessment criteria in marketing plans and advertisements. The study suggests the possibility of moderation of personal innovativeness between quality factors and trust, which could raise new research questions. The study has limitations, including the need for a larger sample size and respondents from different countries to generalize the findings.

Conclusion - The focus of our research was on the adoption of AI-based technologies and the identification of new phenomena that could better describe the adoption of these technologies. Through our findings, we discovered that interaction quality is the primary factor that builds trust in users, which ultimately leads to the intention to use VASs. Our study utilized various theoretical perspectives such as ISSM, SET, and HCI, and highlighted interaction quality as a crucial element in the adoption of new technologies. The results of our research can assist managers in creating more high-quality services and products, as well as attracting more users through the development of better marketing strategies and advertising plans.

(3)

Domain Specific Intelligent Personal Assistant with Bilingual Voice Command Processing.

Author - Saadman Shahid Chowdhury, Atiar Talukdar, Ashik Mahmud

Year - 2018

Abstract - Intelligent Personal Assistants (IPA), like Siri and Alexa, are created to assist their users with simple digital tasks. Here, we propose the steps we have used to develop a voice operated IPA which can process direct commands in two languages: English and Bengali, to perform menial tasks for the users. The speech recognition engine of the IPA is constructed with Sphinx-4, and the language processing is performed by a modified finite state automaton. The IPA also takes advantage of the subject/action structure of commands to reduce the size of the word domain, and utilizes a generalization function to ensure that the language processor can understand multiple languages without undergoing major modification - making this approach suitable when training data is limited.

Introduction and Methodology - An Intelligent Personal Assistant (IPA) is a computer program with Artificial Intelligence (AI) designed to aid its users with their tasks. Modern IPAs can perform a wide variety of tasks ranging from binary to complex tasks. IPAs communicate seamlessly with their users, and modern IPAs are pushing towards Voice User Interfacing (VUI) where the IPA interacts with users only through voice without the use of screens or physical interaction. To achieve this, the IPA needs to listen to human speech, understand what is being implied and perform an action or reply with its own synthesized voice. This paper describes the steps in detail and provides instructions on how to build the ASR and NLP and how to improve the effectiveness of the IPA. The paper also provides pointers on how their work can be further improved upon to create a more effective IPA.

The article presents a methodology for an Intelligent Personal Assistant (IPA) that enables a user to interact with a device using voice commands. The IPA includes an Audio Processing module, Speech Recognition Engine, Natural Language Processing module, and a Generalizer to allow the system to understand multilingual voice commands. The IPA identifies the subject words and action words in a voice command to understand the user's intent better. The system is designed with efficiency in mind, and fewer words are used, resulting in higher success rates. The article explains the flow of information through the IPA and describes each step in detail. The article presents a methodology for building an intelligent personal assistant (IPA) that can interpret spoken commands and perform tasks based on them. The flow of information through the IPA is shown in Figure 2, with steps ranging from audio processing to observable action. Step 2 involves audio processing, where the user speaks into a microphone, and the audio file is converted to mono channels, 16000Hz sampling rate, and 32-bit sample size to standardize audio inputs for the speech recognition engine. Step 3 involves the Sphinx speech recognition engine, which requires a language model, acoustic model, and a dictionary. The dictionary contains the words that the ASR can classify, and the LM contains the phonetic breakdown of those words. The AM contains probabilistic values used to identify phonemes in audio files through hidden Markov models. The IPA is designed for efficiency, identifying a structure in direct verbal commands where there is at least one subject and one action that needs to be performed on that subject. By identifying only subject and action words and discarding the rest, the IPA can deduce what needs to be done with fewer words and less processing. Step 4 involves the generalizer, which allows the IPA to be multilingual. It takes each word as input and outputs a token or null value for each word, with an outputted token being a more "general definition" of the

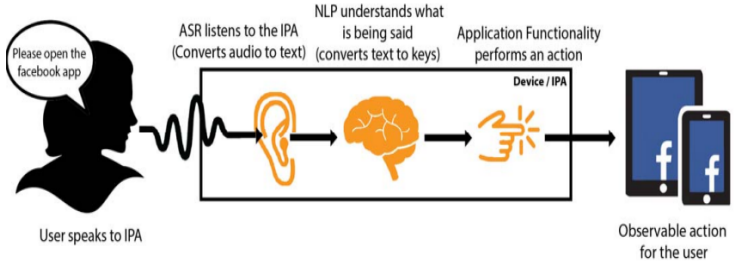


Fig. 1. Concept of IPA

inputted word. The generalizer is important because there are multiple words that imply the same meaning in a certain context.
Overall, the article provides a detailed methodology for building an intelligent personal assistant that can interpret spoken commands efficiently and perform tasks based on them.

Limitations - The Sphinx-4 framework has its own audio noise reduction system, but it is limited in its abilities, and from our experience, attempting to add our own noise cancelling module increases the word error rate of existing acoustic models. The FSA needs to be designed by the developer - hence, when creating an FSA to process more complex commands, the developer needs to ensure that all possible cases are accounted for. However, for simple and direct commands and when datasets are not available, the FSA with Generalizer is a much more cost effective solution than machine learning. As for the Generalizer, two words may have the same spelling but have different meanings, e.g. “bat” can be a playing stick or a flying mammal. This ambiguity is not prevalent in domain specific applications, but may be an issue when scaling up.

Conclusion - The approach specified in this paper is suitable for creating an Intelligent Personal Assistant which is capable of understanding domain specific direct voice commands from users. By selecting only the subject words and action words for training, the size required for the corpus is smaller. The generalizer and FSA allows the creation of an efficient solution to natural language processing when developers do not have access to sufficient training data. Although, this approach has limitations (i.e. it is useful only when the IPA needs to be domain specific), it can be further improved. Some recommended improvements: by using a larger audio/speech corpus for training the CMU Sphinx, the success rate can be increased significantly, the Generalizer can also be designed to be context sensitive to lessen the issue of ambiguity, and the addition of a voice synthesizer after the application functionality step can turn the IPA into a complete VUI.

(4)

A Hybrid Speech Enhancement Algorithm for Voice Assistance Application.

Author - Sajjad Hussain, Abdul Ghafoor Abbasi, Abdul Samad, Abdul Waheed Malik

Year - 2021

Abstract - The paper presents a hybrid speech enhancement algorithm for improving the speech quality in voice assistance applications. The proposed algorithm combines two techniques: spectral subtraction method and Wiener filtering method. The algorithm aims to reduce noise and enhance speech signals in order to improve speech quality. Objective measures such as signal-to-noise ratio (SNR), perceptual evaluation of speech quality (PESQ), and short-time objective intelligibility (STOI) are used to evaluate the effectiveness of the proposed algorithm. Results show that the proposed algorithm outperforms existing methods in terms of speech quality improvement and noise reduction. The research concludes that the proposed hybrid speech enhancement algorithm can significantly enhance the user experience in voice assistance applications.

Introduction and Methodology - The research paper proposes a hybrid speech enhancement algorithm that combines two techniques: spectral subtraction method and Wiener filtering method. The algorithm aims to reduce noise and enhance speech signals in voice assistance applications. The paper explains the methodology of the proposed algorithm and its application in speech enhancement.

Observations and Findings - The effectiveness of the proposed algorithm is evaluated using objective measures such as SNR, PESQ, and STOI. The results show that the proposed algorithm outperforms existing methods in terms of speech

quality improvement and noise reduction. The research paper also compares the proposed algorithm with existing techniques, and the results show that the proposed algorithm performs better in enhancing the speech quality.

Conclusion - The research paper concludes that the proposed hybrid speech enhancement algorithm can significantly improve speech quality in voice assistance applications, thereby enhancing the user experience. The paper emphasizes the importance of speech enhancement algorithms in improving the performance of voice assistance applications. The proposed algorithm can be applied in various fields, including speech recognition, hearing aids, and telecommunication.

(5)

Voice Assistant for visually impaired people

Author -Deepa Gupta, Shikha Mehta, and Shubhi Purwar

Year - 2020

Abstract - The research paper proposes a voice assistant system for visually impaired people to improve their quality of life. The voice assistant system aims to assist visually impaired people in daily life activities, such as navigation, reading, and communication. The system incorporates natural language processing and machine learning algorithms to understand and respond to voice commands.

Introduction and Methodology - The research paper introduces the challenges faced by visually impaired people in their daily lives and highlights the importance of voice assistant systems in improving their quality of life. The paper presents the methodology for designing the voice assistant system, which includes natural language processing, speech recognition, and machine learning algorithms. The system is designed to respond to voice commands and perform tasks such as navigation, reading, and communication.

Observations and Findings - The effectiveness of the voice assistant system is evaluated through user testing with visually impaired people. The results show that the system is effective in assisting visually impaired people in daily life activities and improving their quality of life. The users found the system to be easy to use and helpful in performing tasks such as navigation and reading.

Conclusion - The research paper concludes that the voice assistant system is a valuable tool for visually impaired people in improving their quality of life. The system incorporates natural language processing and machine learning algorithms to understand and respond to voice commands, making it easier for visually impaired people to navigate, read, and communicate. The paper emphasizes the importance of further research in this field to improve the effectiveness of voice assistant systems for visually impaired people.

(6)

A Survey on Voice Assistant Security: Attacks and Countermeasures

Author - Chen Yan, Xiaoyu Ji, Kai Wang

Year - 2022

Abstract - Voice assistants (VA) have become commonplace on a variety of personal devices, including smartphones and smart speakers. Attacks that trick a voice assistant into performing malicious behaviours can pose a significant threat to a user's security, privacy, and even safety as companies build voice assistants with extra functionalities. However, the

literature's various attacks and stand-alone defences frequently lack a systematic perspective, making it difficult for designers to properly identify, understand, and mitigate security threats against voice assistants. To address this issue, this article provides a comprehensive overview of voice assistant attacks and countermeasures. We categorise existing countermeasures based on a broad category of relevant but seemingly unrelated attacks by vulnerable system components and attack methods.

Introduction and Methodology - A voice assistant (VA) receives and executes user voice commands such as making phone calls, playing music, searching for answers, and controlling home appliances. It is not surprising that voice assistants have been widely deployed on smartphones, laptops, smart speakers, vehicles, industrial applications [160], and even military warships [105]. As of 2019, an estimated 3.25 billion voice assistants were in use globally, with the number expected to reach around 8 billion by 2023—nearly one VA per person on average [146]. Attacks that trick a voice assistant into performing malicious behaviours can pose a significant threat to the owner's security, privacy, and even safety as companies rush to build voice assistants with more functionalities.
Gap 1: **Diverse attacks**. Voice assistants are intricate systems composed of numerous software and hardware components. As a result, relevant attacks may involve disparate vulnerabilities, threat models, and methods that appear to be completely unrelated, making it difficult to understand security threats in a systematic manner.
Gap 2: Stand-alone defences. In stand-alone configurations, most existing defences are designed to mitigate only one type of attack. For example, liveness detection is proposed to detect only voice spoofing attacks, whereas adversarial training is intended to resist adversarial example attacks. It is unknown how these defences might apply to voice assistants in a complex adversarial environment, particularly how they compare in terms of implementation cost, usability, and security. Despite the encouraging results presented in numerous articles, it remains difficult for VA designers to select and implement appropriate protection.
Gap 3: Lack of systematic perspectives. Many relevant studies focus on a single VA component, such as speech recognition or speaker verification, rather than the entire system, and in some cases, there is no VA context at all. For example, the majority of attack and defence articles on the security of speech recognition, i.e., the core component of a VA, focus on stand-alone models, many of which have yet to be used in commercial voice assistants. These studies, which may not be indexed with a VA keyword, may, however, apply to voice assistants in the future and will be treated equally.

These gaps have made it difficult to properly identify, comprehend, and mitigate security threats posed by voice assistants.
The methodology is organised as follows: A brief overview of voice assistants. An overview of the attacker's goal, threat model, existing attack methods, and the idealism of a practical attack, as well as an introduction to how this article organises attacks based on vulnerable system components. Then we go into detail about attacks that target the voice assistant's sound-to-audio, audio-to-text, audio-to-identity, and text-to-intent subsystems. Systematising the defence strategies capable of detecting or preventing the aforementioned attacks. Finally we discuss future research directions and make recommendations to voice assistant designers and users and bring this article to a close.
Attack detection: Liveness, Identity. Attack prevention: Hardware-software enhancement, Audio transformation.
Observations and Findings - It is observed that specialised defence methods show good effectiveness and cost for specific attacks whereas all-purpose defence methods can work well for multiple attacks at the same time. Based on an overall comparison of defence methods, they are classified into three categories:
Specialised, all-purpose, and complementary.
Specialised eg: — Voiceprint verification against Normal Speech attacks.
All-purpose eg: — liveness and identity-based detection methods
Complementary — Other methods

Conclusion - We systematised on the five types of attack methods, i.e., normal speech, voice spoofing, unintelligible speech, adversarial example, inaudible signal, and malicious skill, based on the vulnerable VA subsystem. They divided existing countermeasures into two categories, i.e., detection and prevention, and systematised them by the shared defensive strategies to compare their applicability by the implementation cost, usability, and security.
Identify following trends: Adversarial Inaudible Signal Attacks, Adversarial Examples against both Audio-to-Text and Audio-to-Identity, Robust Defence.

(7)

Voice Assistance in 2019

Author -Robert Dale

Year - 2019

Abstract - The end of the calendar year always seems like a good time to take a breather and reflect on what's happened in the previous year, and this is as true in the world of commercial NLP as it is in any other. 2019 has been a busy year for voice assistance in particular, thanks to the emphasis placed on this area by all of the major technology players. As a result, we'd like to take this opportunity to go over a few key themes that have defined recent developments in the commercialization of voice technology.

Introduction and Methodology - In this post, I reflect on what has been in the news over the last year and identify what I see as the major themes that have defined voice developments in 2019.
1. A clash of titans: Amazon Alexa vs. Google Assistant - If you want to buy a smart speaker, you'll almost certainly have to choose between an Amazon Echo and a Google Home. Although there are a few other players in the market—and we'll get to those later—the main contenders at the moment are devices powered by Amazon's Alexa, which celebrated its fifth birthday in November 2019, and Google's Assistant, which celebrated its third birthday in October.
2. Additional voice assistants - Of course, there are other players in the voice assistant market. In fact, Chinese manufacturers, about whom little is heard in the West, had a good showing in 2019Q3: Alibaba and Baidu narrowly outsold Google in smart speakers, shipping 3.9 and 3.7 million units, respectively, and Xiaomi trailed Google by 3.4 million units.
Other voice assistants, on the other hand, have been struggling. Cortana, Microsoft's digital assistant introduced in 2014 to compete with Siri as a digital assistant for the now-defunct Windows Phone, had become a key feature of the Windows 10 interface by the beginning of 2019.
3. Conversational ability - All of this competition has inevitably contributed to a steady stream of advancements in the technology underlying voice assistants. Google Duplex, a hyper-realistic appointment-making voice dialogue app that the company demoed at Google I/O and piloted in New York, Atlanta, Phoenix, and San Francisco towards the end of that year, was the big news in 2018. This year saw the gradual implementation of that technology: by mid-year, Duplex was available for restaurant reservations in 43 US states, and a New Zealand trialFootnotet was proposed by the end of the year.
4. Voice app development tools - The complexity of the conversation in the Google Duplex demo is in a completely different league than what happens when you ask your Google Home about the weather. As a small step towards closing the gap, all of the major vendors have added features to their developer platforms that enable extended conversations, allowing users to go beyond simple one-time question-and-answer dialogues.
5. Speech Synthesis - Several vendors announced incremental improvements in speech recognition performance throughout the year, but these are not always obvious to the end user. Speech synthesis advancements are much more visible (audible?).Earlier this year, Amazon announced the general availability of its newsreader style in the Amazon Polly

text-to-speech service, Footnoteak, which provides pretty impressive intonation: the sample in this VentureBeat piece is worth checking out. And, at least in the United States, you can now ask Alexa to speak slower or faster.
6. Confidentiality - In terms of AI applications in general, 2019 was the year in which a number of ethical concerns were raised. The big issue for voice this year was data privacy.We've been concerned about listening devices since we first invited them into our homes, but our initial concerns were alleviated by vendor assurances that our smart speakers only began listening after the appropriate wake word was uttered.
7. The prevalence of voice - So it's been a long year, and it appears that voice is almost everywhere. To reinforce that point, here are 3 of the 12 new things you can do with your voice that debuted in 2019.
(1) Speak with more kitchen machines.
(2) Drive-thru
(3) Smart-sprinklers

Observations and Findings - It is observed that the technology related to voice assistants is getting better and better with google and amazon leading the charge.Alexa beats google assistant in terms of sales and privacy while google still has the upper hand in terms of accuracy of answers.

Conclusion - Amazon Alexa vs Google Assistant, conversational ability in which Google's Duplex is becoming more and more advanced, tools for building voice apps like Alexa conversations, improvements in Google's Cloud Text-to-Speech, smart sprinklers which have had an upgrade as well are all major voice related strides taken in the year of 2019. Overall, the VA domain was thoroughly explored for the year of 2019.
In the future, research can be done into applications for jobs using smart assistants.

(8)

Humanising voice assistant: The impact of voice assistant personality on consumers' attitudes and behaviours

Author - Atieh Poushneh

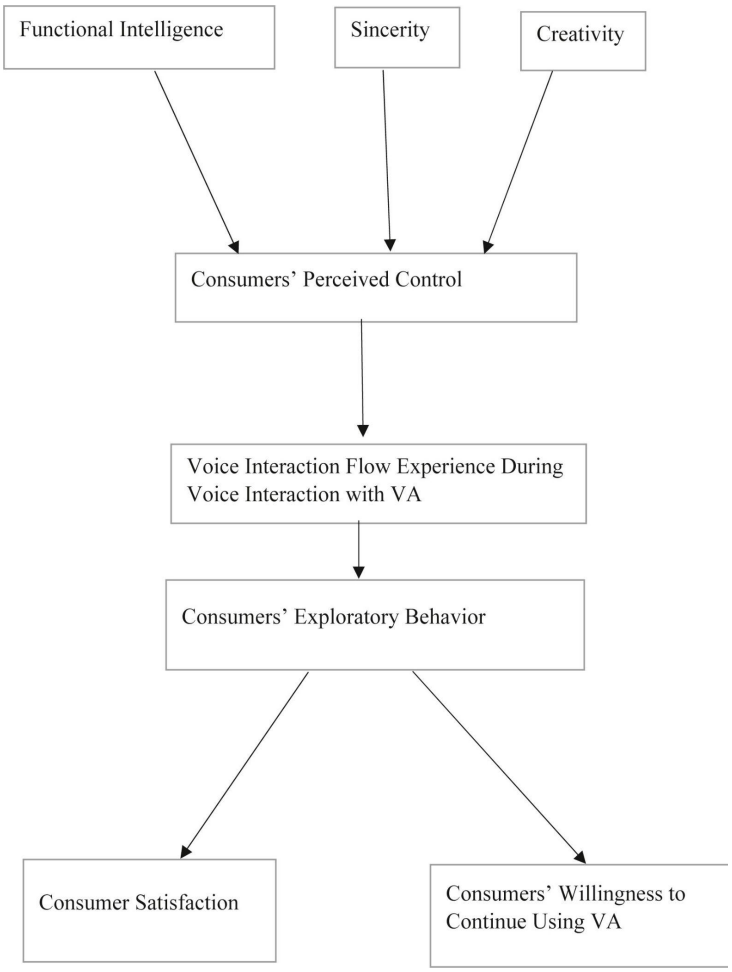
Year - 2020

Abstract - A voice assistant (VA), a type of voice-enabled artificial intelligence, is no longer just a movie character. Voice is now embedded in a wide range of products, including smartphones (mobile applications) and smart speakers in consumers' homes. Furthermore, voice assistants are becoming more and more common in our daily lives. While human personalities shape how we interact with the world, voice assistant personalities can also have an impact on how we interact with our surroundings on a daily basis. This study identifies seven voice assistant personality traits (VAP) of three widely used mobile applications: Cortana from Microsoft, Assistant from Google, and Alexa from Amazon. This study applies and extends flow theory to investigate why VAP has the effects it does and what aspects of VAP drive the voice interaction flow experience that can influence consumers' attitudes and behavioural intentions. According to our findings, voice interaction with a virtual assistant that incorporates functional intelligence, sincerity, and creativity enables consumers to take control of their voice interactions with the VA, focus on their voice interaction, and engage in exploratory behaviour. Consumer satisfaction and willingness to continue using voice assistants are influenced by consumers' exploratory behaviour.

Introduction and Methodology - Voice assistants (VA) are a type of artificial intelligence that uses voice commands (AI). AI refers to the ability of digital interfaces to display some level of intelligence, or the ability of algorithms to mimic

intelligent human behaviour. Although artificial intelligence refers to "cognitive" functions associated with the human mind, such as problem solving and learning.

Apple's Siri, Amazon's Alexa, Google Assistant, Microsoft Cortana, and Amazon's Echo are among smart speaker offerings, as are Google's Home and Apple's Home. In any form, virtual assistants (VA) are revolutionising consumer consumption culture and becoming an increasingly important part of consumers' social lives. Users can use such virtual assistants to navigate, listen to music, send text messages, control smart home devices, make a phone call, order food, order an Uber ride or pizza, and so on. According to National Public Radio and Edison Research, 21% of Americans (53 million people) own smart speakers, a significant increase from the 14 million who did so in 2018. Huffman, Google Assistant's Vice President, announced that the Google Assistant mobile application has been downloaded on 500 million devices. Google Assistant is compatible with over 1000 different smart machines, including dishwashers, ovens, and light bulbs (Wiggers, 2019).



Algorithms are being developed by developers to provide VA, social characteristics, and specific personalities. A recent study documented people's reactions to a robot asking them not to turn it off (Horstmann et al., 2018), demonstrating that people respond socially to robots that mimic human behaviour. It's difficult to turn off a robot that begs you not to. Amazon's AI developers are currently developing applications to give Alexa a distinct personality, as well as to make her more "conversational," recall more information, and engage in longer conversations (Roettgers, 2019). (Rubin, 2017). If you ask Alexa how she is, she will tell you that she is "feeling pretty studious since the holidays have arrived and she has been learning some fun Kwanzaa facts." Alexa's abilities extend beyond simple tasks to provide users with her opinions. However, according to Amazon Senior Vice President Dave Limp, we still don't know how much personality consumers attribute to voice assistants.

Convenience sampling was used to recruit participants. All VA personality traits and consumer satisfaction were measured using a 7-point Likert scale with the anchors being "strongly disagree" and "strongly agree".

Observations and Findings - It is observed that VA personality traits drive voice interaction flow experience resulting in satisfactory experience through voice interaction.

Results indicate that functional intelligence, sincerity, aesthetic appeal of VA mobile interface, creativity, and protective quality varied widely across groups;

However, sociability and emotional intelligence held steady among the three applications.

Conclusion - VA manifesting functional intelligence, sincerity, and creativity traits can enhance consumers' perceived control during voice interactions, especially when the VA offers effective, efficient, sincere, honest, and current

information. As consumers learn through their voice interactions with VAs, they become more confident and more willing to seek informationFuture research should evaluate the effect of voice recognition, social presence and social image on consumers' cognitive load and add other smart devices like Apple Home for comparison purposes.

(9)

Investigation and development of the intelligent voice assistant for the Internet of Things using machine learning

Author - E. V. Polyakov, M. S. Mazhanov, A. Y. Rolich

Year - 2018

Abstract - The appearance and widespread adoption of the Internet of Things has facilitated the active use of artificial intelligence technologies in human life (IoT). Autonomous devices are becoming smarter in how they interact with humans and with one another. New capabilities lead to the development of various systems for integrating smart things into Internet of Things Social Networks. One of the important trends in artificial intelligence is the technology for recognising human natural language. New insights in this area may lead to new forms of natural human-machine interaction, in which the machine learns to understand human language and adjusts and interacts with it. Voice assistant is one such tool that can be integrated into a variety of other intelligent systems. The principles of voice assistant operation are described in this paper, as are the main shortcomings and limitations. The method for creating a local voice assistant without using cloud services is described, which allows for the future applicability of such devices to be significantly expanded.

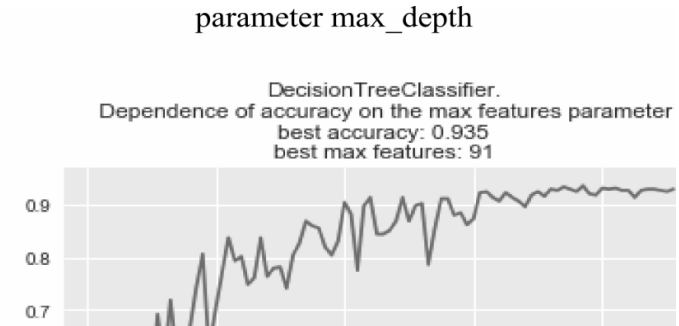
Introduction and Methodology - Today, the development of artificial intelligence (AI) systems capable of organising natural human-machine interaction (via voice, communication, gestures, facial expressions, and so on) is gaining traction. One of the most studied and popular was the direction of interaction, which was based on the machine's understanding of natural human language by the machine. It is no longer the case that a human learns to communicate with a machine, but rather that a machine learns to communicate with a human, exploring his actions, habits, and behaviour in order to become his personalised assistant.

Work on developing and improving such personalised assistants has been ongoing for quite some time. These systems are constantly evolving, have moved beyond personal computers, and have firmly established themselves in a variety of mobile devices and gadgets. Siri from Apple, Amazon Echo, which responds to the name Alex from Amazon, Cortana from Microsoft, Google Assistant from Google, and the recently launched intelligent assistant from Yandex are among the most popular voice assistants.

Sections I and II provide an overview of the architecture and construction of voice assistants. Section III illustrates the work of a voice assistant based on Alice, a Yandex product. Section IV discusses the flaws of existing voice assistants and how to fix them. Tools are required for the development of such systems. Section V describes the methods for developing and training an assistant using various machine learning algorithms, as well as a comparative evaluation of algorithm learning ability. The primary goal of this work is to create a local voice assistant that does not rely on various cloud technologies and services, allowing it to be used to solve a variety of specific problems.

PocketSphinx tools are used for automatic voice recognition, and the Festival engine is used to generate the voice.

Machine Learning algorithm used: Single classifier using MultiOutputClassifier object in Scikit-Learn, python.



Observations and Findings - It is observed that for the presented data set (about 450 samples), the best result shows the algorithm of "Trees of

Solutions" with an accuracy of 93%. Besides, the study revealed the creation and the use of voice assistants is not limited only to cloud services. Summarising, we can conclude that the best results of learning on a small amount of data showed the algorithm "Decision tree" with 93% of correct answers, "Polynomial Bayesian Classifier" with the correctness - 81% and the lowest quality in the "k-nearest neighbours" method with a correctness of 73%.

Conclusion - The work utilised the principles of voice assistants that are currently available on the market. The main flaws in relevant research were identified. A method for addressing these flaws was proposed. A voice assistant is being developed and trained. In addition, the learning ability of the algorithms for recognising intentions was evaluated. As a result, the best result for the presented data set (about 450 samples) shows the "Trees of Solutions" algorithm with an accuracy of 93%. Furthermore, the study revealed that the development and use of voice assistants is not limited to cloud services.
Furthermore, the use of local systems allows for the expansion of the range of tasks in which they can be used in IoT and IIoT systems, smart home systems, healthcare, security, and systems with a high level of confidentiality, where the use of cloud technologies can be difficult. Different tools can be used for generation of automatic voice recognition to pinpoint subtle differences in the intelligence of the voice assistant.

(10)

A Smart Personal AI Assistant for Visually Impaired People

Author - Shubham Melvin Felix, Sumer Kumar, A. Veeramuthu

Year - 2018

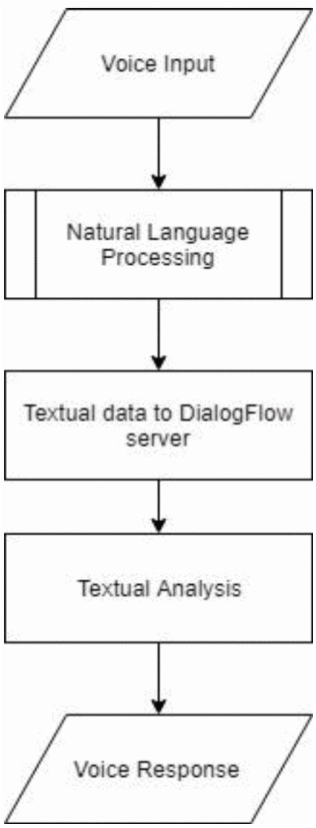
Abstract - In today's advanced hi-tech world, the need for independent living is recognised in the case of visually impaired people who face the main problem of social constraint. They suffer in strange environments with no human assistance. Because visual information is the foundation of most tasks, visually impaired people are at a disadvantage because necessary information about their surroundings is unavailable. With recent advancements in inclusive technology, it is now possible to expand the assistance provided to people with visual impairment. This project proposes using Artificial Intelligence, Machine Learning, Image and Text Recognition to assist people who are blind or visually impaired. The concept is realised through an Android mobile app that focuses on voice assistant, image recognition, currency recognition, e-book, chat bot, and other features. The app can recognise objects in the environment using voice commands and can perform text analysis to recognise text in a hard copy document. It will be an efficient way for blind people to interact with the environment and utilise technology's facilities.

Introduction and Methodology - Visual impairment is a significant disability that visually challenged people face. A person who cannot see will never feel the same emotions as someone who can see the world. This visibility issue is a black dot that affects billions of people worldwide. Our goal is to eliminate this black dot using Artificial Intelligence and Machine Learning.
Visual impairment has serious consequences for certain visual function abilities:
The daily routine (that requires a vision at an average distance)
iTalking, reading, and writing (which requires a precise vision and average distance)
Area and displacement estimation (which require a far vision)
The tracking of an activity necessitates a prolonged period of optical observation.

The current system employs speech synthesis to read e-books for the visually impaired via a mobile application and converts the books' document/soft copy to speech via natural language and Text-to-Speech.

The existing system has a major flaw in that it only supports one language (English) and is incompatible with other languages. It does not function offline and requires an internet connection to provide feedback/response.

The proposed system, which is entirely based on voice commands, employs Artificial Intelligence to assist visually impaired people. It also recognises images in photographs or uses a camera to recognise objects and describe them in audio, and it has a chat bot for light and friendly conversations.



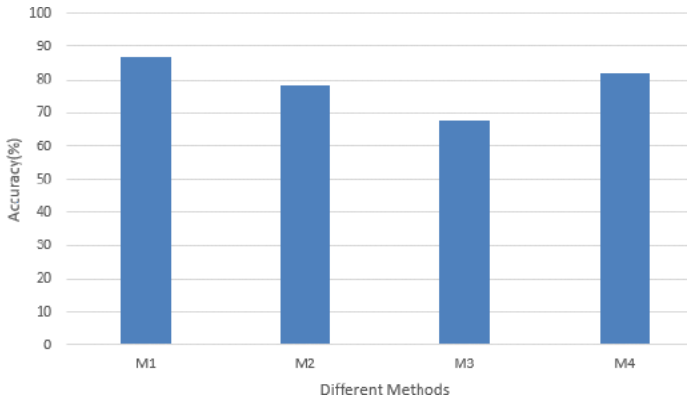
The remaining sections are arranged as follows: Section II discussed various related works done in this field. This section III contains the proposed work, including the problem statement, complete system architecture, and algorithm description. Section IV discussed the results and the work that was completed. Finally, in Section V, the work was completed and future directions were provided.

DialogFlow platform is used to train a chat-bot integrated with Cloud APIs of Google, so that a voice input by a blind user can generate appropriate voice response.

Observations and Findings - It is observed that the proposed system works better and incorporates more domains than existing systems.The comparison done between object recognition, landmark detection, textual analysis, interactive, voice input as well as response and multilingual properties shows that the proposed system acquires more features than others do.

The table compares features between different research papers according to the accuracy of features which is shown in "Table I". Object recognition, landmark detection, textual analysis, interactive, voice input, response, and multilingual properties were compared. We can conclude from the comparison table that our proposed system has more features than others.

Features	M 1	M 2	M 3	M 4
Object Recognition	✓	✓	✓	✗
Landmark Detection	✓	✗	✗	✗
Textual Analysis	✓	✗	✗	✓
Interactive	✓	✗	✗	✓
Voice Input & Response	✓	✗	✓	✗
Multilingual	✗	✗	✗	✗



The bar graph below compares the accuracy of our paper to that of other papers. The graph shows that our proposed system is more accurate than the other proposed systems. Our proposed system M1 has an accuracy of 80 to 90, which is the highest among all.

Conclusion - Artificial intelligence and machine learning are two of the most rapidly evolving technologies. These technologies are critical to the development of the IT sector. We attempted to use these technologies for visually impaired

people so that they could live an independent and normal life. The pleasant conversation with the bot Object and environment image recognition. Currency recognition to facilitate payment. Text recognition and analysis. If the proposed system is completed, it will provide a better assistant to visually impaired people.

In the coming days, our proposed system will be implemented in multilingual applications, allowing users to use the application in their native language without difficulty. Furthermore, our proposed system is compatible with IoT. In the future, our proposed system will be much better at interpreting textual descriptions. Image recognition can be improved by providing more information about the image captured by the camera. This system can be improved by incorporating currency recognition features [12]. The existing methodology for image and currency recognition can be improved.

AI and ML technologies are used for the visually challenged people so that they can also live an independent and normal life. The friendly chat with the bot Image recognition of the objects and surroundings. Currency recognition and Text recognition for daily chores.

In the future, the system can be applied in multilingual applications. In addition, the proposed system can be deployed with the IoT. The proposed features can be improved as well.

Literature Review Summary:

Sr No	Author and year of Publications	Paper title	Observations and Remarks	Comments
1	Subhash S, Prajwal Srivasta, Siddesh S, Ullas A, Santhosh B 2020	Artificial Intelligence based Voice Assistant.	It can behave like a personal bot for typical search, perform tasks under permissions,play or download a song, take screenshots, highlight important locations specified by the user on google maps, give live news around the world, give security checks. If a person is in danger, it can be used to send the user's location to the police or close relatives by giving the command “I’m in danger”.	Reducing redundant noise, using sensors to command, using acoustic analysis to process and gTTS to display appropriate results.
2		AI-Based Voice Assistant Systems: Evaluating from the Interaction and Trust Perspectives.		
3		Domain Specific Intelligent Personal Assistant with Bilingual Voice Command Processing.		
4		A Hybrid Speech Enhancement Algorithm for Voice Assistance Application.		
5		Voice Assistance for Visually Impaired People.		
6		A Survey on Voice Assistant Security: Attacks and Countermeasures.		
7		Voice assistance in 2019.		
8		Humanizing voice assistant: The impact of voice assistant personality on consumers’ attitudes and		

		behaviours.		
9		Investigation and development of the intelligent voice assistant for the Internet of Things using machine learning.		
10		A Smart Personal AI Assistant for Visually Impaired People.		
11		Dynamic AI based Email Voice Assistant for Web Services.		

Title of paper	**Introduction and Methodology**	**Observations and Result**	**Conclusion and Future scope**
Voice Assistance for Visually Impaired People.	This paper describes a device for visually impaired people, to read environmental messages, words, letters, newspapers and so on. Earlier, people with less or no eye sight used braille type of paper to read messages or	Blind people face many problems in day to day activities. Nowadays, textual information is everywhere around and its important to integrate these messages so that they can lead a normal life with society. A system can be defined in such a way that those people can communicate or listen to it in the language of their origin. It uses cameras, OCR and TTS to communicate and perceive. The principle of the device is to have an interactive system to	The Paper helped blind or partially blind people lead a good social life in the community. They required less assistance in getting and comprehending environmental messages. Future is vast, such as the detection of their daily path, google assistance and gps tracker can also help in performing tasks or daily activities like guiding, getting

	required assistance They used OCR (Optical Character Recognition) sensors, TTS (Text to Speech) modules, Raspberry pie as motherboard, Digital Camera for scanning with adapted features.	capture the image.Basic Problem which arises was while taking the picture, object-to-device distance with additions to environmental factors such as brightness played an important factor. For each snapshot taken , the system assists them in taking the picture by guiding them or indicating the quality of sight. It conveys appropriate messages such as “The image is fuzzy” , “bring camera closer”, “low brightness”, etc.	information about the destination. All of this can be integrated into raspberry pie which will make it a fully fledged assistant for the people in need.
A Survey on Voice Assistant Security: Attacks and Countermeasures	Attack detection: 1)Liveness 2)Identity Attack prevention: 1)Hardware-software enhancement 2)Audio transformation	It is observed that specialized defense methods show good effectiveness and cost for specific attacks whereas all-purpose defense methods can work well for multiple attacks at the same time. Based on an overall comparison of defence methods, they are classified into three categories: Specialised, all-purpose, and complementary. Specialised eg: — Voiceprint verification against Normal Speech attacks. All-purpose eg: — liveness and identity-based detection methods Complementary — Other methods	They systematised on the five types of attack methods, i.e., normal speech, voice spoofing, unintelligible speech, adversarial example, inaudible signal, and malicious skill, based on the vulnerable VA subsystem. They divided existing countermeasures into two categories, i.e., detection and prevention, and systematised them by the shared defensive strategies to compare their applicability by the implementation cost, usability, and security. Identify following trends: Adversarial Inaudible Signal Attacks, Adversarial Examples against both Audio-to-Text and Audio-to-Identity, Robust Defence.
Voice assistance in 2019	A comprehensive survey of voice assistants by analysing device counts, sales and accuracy of answers, tools for building them, conversational ability and privacy using data available online.	It is observed that the technology related to voice assistants is getting better and better with google and amazon leading the charge.Alexa beats google assistant in terms of sales and privacy while google still has the upper hand in terms of accuracy of answers	Amazon Alexa vs Google Assistant conversational ability in which Google’s Duplex is becoming more and more advanced, tools for building voice apps like Alexa conversations, improvements in Google’s Cloud Text-to-Speech, smart sprinklers which have had an upgrade as well are all major voice related strides taken in the year of 2019. In the future, research can be done into applications for jobs using smart assistants.
Humanizing voice assistant: The impact of voice assistant personality on consumers’ attitudes and behaviours	Convenience sampling was used to recruit participants. All VA personality traits and consumer satisfaction were measured using a 7-point Likert scale with	It is observed that VA personality traits drive voice interaction flow experience resulting in satisfactory experience through voice interaction. Results indicate that functional intelligence, sincerity, aesthetic appeal of VA mobile interface, creativity, and protective quality	VA manifesting functional intelligence, sincerity, and creativity traits can enhance consumers’ perceived control during voice interactions, especially when the VA offers effective, efficient, sincere, honest, and current information. As

	the anchors being "strongly disagree" and "strongly agree".	varied widely across groups; However, sociability and emotional intelligence held steady among the three applications.	consumers learn through their voice interactions with VAs, they become more confident and more willing to seek informationFuture research should evaluate the effect of voice recognition, social presence and social image on consumers' cognitive load and add other smart devices like Apple Home for comparison purposes. .
Investigation and development of the intelligent voice assistant for the Internet of Things using machine learning	PocketSphinx tools are used for automatic voice recognition, and the Festival engine is used to generate the voice. Machine Learning algorithm used: Single classifier using MultiOutputClassifier object in Scikit-Learn, python.	It is observed that for the presented data set (about 450 samples), the best result shows the algorithm of "Trees of Solutions" with an accuracy of 93%. Besides, the study revealed the creation and the use of voice assistants is not limited only to cloud services. Summarising, we can conclude that the best results of learning on a small amount of data showed the algorithm "Decision tree" with 93% of correct answers, "Polynomial Bayesian Classifier" with the correctness - 81% and the lowest quality in the "k-nearest neighbours" method with a correctness of 73%.	Different tools can be used for generation of automatic voice recognition to pinpoint subtle differences in the intelligence of the voice assistant.
A Smart Personal AI Assistant for Visually Impaired People	DialogFlow platform is used to train a chat-bot integrated with Cloud APIs of Google, so that a voice input by a blind user can generate appropriate voice response.	It is observed that the proposed system works better and incorporates more domains than existing systems.The comparison done between object recognition, landmark detection, textual analysis, interactive, voice input as well as response and multilingual properties shows that the proposed system acquires more features than others do.	AI and ML technologies are used for the visually challenged people so that they can also live an independent and normal life. The friendly chat with the bot Image recognition of the objects and surroundings. Currency recognition and Text recognition for daily chores. In the future, the system can be applied in multilingual applications. In addition, the proposed system can be deployed with the IoT. The proposed features can be improved as well
Dynamic AI based Email Voice Assistant for Web Services			

Conclusion : We have successfully studied 15 Research Papers on the topic “Artificial Intelligence for Voice Assistance”.