iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** V **Month of publication:** May 2024

DOI: https://doi.org/10.22214/ijraset.2024.61842

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

Artificial Intelligence: A Quick Overview of Speech Recognition System

Ms. Jatinder Kaur[1], Ms. Swati Sharma[2], Dr. Varun Tiwari[3], Dr. Sanjay Kumar Singh[4]
[1]*Assistant Professor, Don Bosco Institute of Technology, Okhla Road, New Delhi, India*
[2]*Assistant Professor, Bosco Technical Training Society, Okhla Road, New Delhi, India*
[3, 4]*Associate Professor, Don Bosco Institute of Technology, Okhla Road, New Delhi, India*

Abstract: ***The study of intelligent behaviour is known as artificial intelligence. Wagner said, "We build intelligent machines in science." By this, he implied that the primary goal is to teach these sophisticated devices, such as computers, to examine and comprehend human conduct. numerous believed that human intelligence might be understood through the construction of several programmes, while other researchers believed that numerous basics should be required in order to accomplish the intended objective. We can carry out or replicate any type of task we conduct using a computer. Therefore, computers are the ideal devices for developing artificial intelligence. In the hopes that they will outperform computers in several other areas, some researchers created additional computing devices. Billions of dollars were invested in creating new devices that surpassed the speed of the computer used to simulate software. However, the computers and the code both need to be extremely quick. As you learn about artificial intelligence, you may wonder, what really is intelligence? How is intelligence measured? Turn become something quite significant. In this case, the programme that evaluates its surroundings acts as the intelligent agent. Modern digital computers and human minds are comparable to one other in ways such as symbolic information processing systems. Both use symbolic data as input, change it in accordance with predetermined guidelines, and then use the results to solve issues. In order to mechanically replicate intelligent human activity, artificial intelligence researchers have programmed some algorithms to recognise intelligent human behaviour. Human conduct in language processing, chess, and other games, as well as medical diagnostics, may be used to monitor this activity. Voice recognition is a key component of AI. A certain voice may be recognised using voice recognition technology. The cornerstone for speaker identification is voice signals. Voice targeting is applicable to a wide range of applications, including voice mail, database access, phone banking, and phone purchasing. The ability to enter one's voice for verification is among the most potent uses of voice recognition for security. The fundamental means of interpersonal communication is speech. The technique of translating voice sounds into appropriate text is known as speech recognition. Over the past few years, speech recognition technology has advanced significantly. Nonetheless, there are other significant study obstacles, such as variations in speaker and language, ambient sound, word size, etc. This study aims to give a comprehensive overview of speech acceptance by summarising the several approaches utilised in the standard speech system and describing the numerous processes involved.***

Keywords: Artificial Intelligence, intelligent, speech recognition, modelling, speech processing, training and assessment.

I. INTRODUCTION

These days, AI technology is growing rapidly. Research on this issue is quite popular since advancements in a variety of fields have benefited society from the 1990s to the late 20s. Artificial intelligence is developing at a rapid pace, which has significant effects on both the economy and society at large. Youth with skill will have more employment as a result of this. Artificial intelligence (AI) is being used in a wide range of industries, including manufacturing, healthcare, the military, and automotive. This improves every element of our everyday existence.

According to the paper's data, India is regarded as a major participant in AI research and development. In terms of scholarly publications connected to AI, it ranks fourth, while in terms of AI patents, it ranks first. India has optimism because artificial intelligence (AI) is accelerating growth and removing long-standing barriers like bureaucracy and inadequate infrastructure.

©IJRASET: All Rights are Reserved

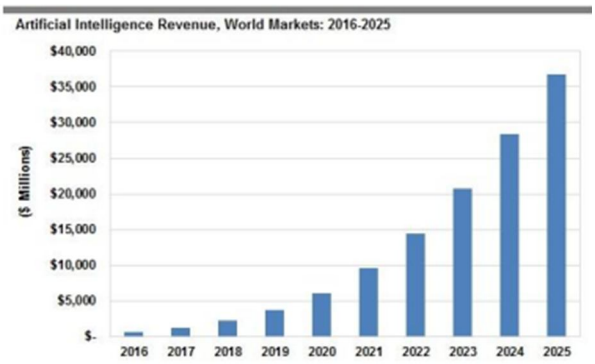


Fig 1: AI market from 2021 to 2025

From 2021 to 2025, the worldwide AI market is projected to expand at a compound yearly growth rate of 40.2%, reaching USD 997.77 billion. Voice is a fundamental, widely used, and efficient form of interpersonal communication that is a key component of artificial intelligence. These days, voice technology can be used for a small but impressive number of purposes. Thanks to technology, robots can now react to human voices correctly and consistently and offer beneficial services. People will like such programme since computer communication is faster than typing on a keyboard. This may be accomplished by creating a computer-to-speech voice recognition system, which enables the computer to translate voice commands and make phone calls from text. Voice recognition system: process-to-text is the process of translating a collection of words from an auditory signal captured using a microphone. The technique of Automatic spoken Recognition (ASR), sometimes referred to as Speech Recognition, fundamentally modifies spoken signals. Developing voice recognition techniques is the goal of the speech recognition platform. Scale and power constraints restricted the early computer programmes. However, the subject of default voice recognition is already seeing a shift in computer technology. Thanks to advancements in computer technology, voice recognition significant details are now easily retained. The languages in which voice recognition systems have evolved are quite limited. As a result, the construction of native language phrases involves several aspects. In several domains, including automated telephone processing on the phone network, data entry, and voice calling—as well as questions based on repurposed travel and reservation data and natural language translators—automatic expression recognition has reduced the need for human intervention. This essay focuses on the fundamental components of voice recognition systems, technological advancements, and issues with automatic speech recognition.

II. LITERATURE REVIEW

In order to control the movement of a mobile robot, Thiang et al.
(2011) demonstrated voice recognition utilising artificial neural networks (ANNs) and linear predictive coding (LPC). LPC and ANN performed the extraction after immediately sampling the input signals from the microphone. A speaker-independent, isolated voice recognition system for the Tamil language was proposed by Ms. Vimala C. and Dr. V. Radha (2012). The use of HMM in feature extraction, acoustic model, pronunciation dictionary, and language model yielded 88% accuracy in 2500 words.
Cini Kurian and Kannan Balakrishnan (2012) discovered the creation and assessment of several acoustic models for continuous speech recognition in Malayalam. There are 21 speakers in the database, 10 of whom are men and 11 of them are women.
In 2013, Suma Swamy and colleagues presented an effective voice recognition system that utilised Mel Frequency Cepstrum Coefficients (MFCC), Vector Quantization (VQ), and HMM, achieving 98% accuracy in speech recognition. Five words are pronounced 10 times by four speakers in the database.
Using the Hidden Markov Model Toolkit (HTK), Annu Choudhary et al. (2013) presented an automated audio recognition system for isolated and linked Hindi language words. The recognition system obtained 95% accuracy in isolated words and 90% accuracy in linked words using the Hindi words that were extracted from the dataset via MFCC. HTK was suggested by Preeti Saini et al. (2013) for Hindi automated speech recognition.
Using isolated words and 10 states in HMM topology, 96.61% of the speech was recognised. (2013) introduced an automated voice recognition method for Bangla words by Md. Akkas Ali et al. Gaussian Mixture Model (GMM) and Linear Predictive Coding (LPC) were used for feature extraction. A total of 1000 recordings of 100 words yielded an accuracy of 84%. In 2014, Maya Money Kumar and colleagues created a speech recognition system for Malayalam word identification.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

The suggested study used HMM on MFCC for feature extraction and syllable-based segmentation. Sanskrit speech was first introduced by Jitendra Singh Pokhariya and Dr. Sanjay Mathur in 2014.

III. HISTORY OF ARTIFICIAL INTELLIGENCE

The technology of artificial intelligence is quite old. Ancient Greek and Egyptian mythology have brief accounts of mechanised human beings. Between 380 BC and 1900 AD, several mathematicians studied mechanical courses, calculators, and integer systems; these developments eventually led to the concept of mechanised human acknowledgment in non-living beings. After the turn of the 20th century, the rate of invention in artificial intelligence increased significantly. Makoto Nishimura, a Japanese scientist and professor, created the nation's first robot in 1929. John Vincent Atanasoff and his colleague created a computer in 1939 that was capable of solving 29 linear equations at once. Many AI-related research projects were undertaken in the 1950s. A scientist by the name of Claude Shannon created a chess-playing computer. The first computer was created in 1960 and operated on a General Motors production line. A movie about a robot that can communicate in seven different ways was released in 1977. A chess machine created by IBM in 1997 defeated the current world champion. A dog robot that learns from its surroundings was produced by Sony in 1999. AI becomes increasingly significant in human daily lives starting in 2010. A few excellent examples of it include Siri, Alexa, and Google Assistant, which interpret sounds and translate them into text using voice recognition technology.

IV. ABOUT VOICE RECOGNITION SYSTEM

There are essentially two approaches to the speech recognition system.

1) To identify the Speaker

2) Speaker authentication

Speaker identification is the technique of recognising the voice of a speech delivered to a specific set of speakers. A speaker whose vibrations are at their maximum is the same as the voice that has been served; this speaker is qualified for a new entry into the database because of their unparalleled voice attributes. Open-set mode and Open-set mode are known sets of sounds encoded in two edges. The speaker is not required to be a part of a group of well-known speakers while using the open-set method. This is applied in cases of particular criminal action from which it originates. In close-range mode, the speaker can identify several suspects since their voice is among the other recognisable voices that are already stored in the database. This technique is used for security purposes to verify an authorised person's identity using biometrics. On the other hand, the process of approving or disapproving a speaker's assertion is known as speaker authentication. It is employed to support someone's appeal for veracity. The process of verifying the authenticity of a voice from a set of speakers is commonly known as the "open set mode" in speaker verification. The most crucial component of speech recognition software is the permission of a certain service, which is emphasised in the identity verification system of every speaker. Text and independent text recognition are the subjects of another segment of the speech recognition technique. This is the distinction that the speaker is referring to in the text. The text saved during training is referred to as the Text Recognition Programme if the text uttered by the speaker is the same as the text stored. Conversely, though, if any Voice independent text programme refers to the spoken text that is provided at random by the speaker using voice identification. Consequently, there are three indicators of a trustworthy speech recognition system: text dependent and independent, open and closed sets, voice recognition, and voice verification. A microphone was used to record speech and convert it into an electrical signal. The purpose of the computer's sound card Transform a signal from analogue to digital. This voice signal can be stored and played by the sound card.

The components of a typical speech recognition system are as follows.

- Signal pre-processing
- Facility extraction
- Language model
- decoder
- Speech recognition

V. TYPES OF SPEECH

The categories of speech recognition are based on the kind of words they can comprehend. They fall under one of the following categories:

1) *Remote Name:* Remote Identifier occasionally demands that every word said by the bot be quiet (have no sound signal).

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

2) *Connected Name:* Same as before, but with a different name that permits several expressions to coexist "has a space, at minimum, between them.
3) *Continuous Speech:* this feature lets people talk normally while the computer chooses the information.
4) *Sound Speech:* This type of speech is incoherent and noisy.

Various types of speech recognition systems exist, contingent upon the word types that need to be monitored. The differences between these different categories are as follows:

a) *Isolated Words:* Individuals studying single words frequently discover that every phrase is serene on both sides of the example window. There are often two listening/voice listening sections in these programmes, when speakers must wait to speak during midnight addresses. Speech signals are processed in the silence that occurs between sentences.
b) *Linked names:* Linked words have a little pause difference between them but otherwise function similarly to single words.
c) *Continuous Words:* A genuine speaking style is required for continuous speech recognition. The challenging design persisted because it searched for precise methods to define speech boundaries.
d) *Default Names:* Nonverbal, nonverbal, and erroneous statements that are challenging to interpret are covered by default expressions. The ASR programme in this part addresses a number of topics, including grouping words like "mask" and "ahs" together.

VI. OVERVIEW OF THE ADVANCEMENTS IN VOICE RECOGNITION TECHNOLOGY

An overview of the advancements in voice recognition technology is given in this section. A succinct review of some techniques that have been employed to enhance the recognition process is also included. According to a poll, the IT industry has made considerable strides in the recent several years.

1) In the 1950s, several researchers attempted to investigate the fundamental acoustic-phonetic notion, which led to the first attempts to construct automatic conversation recognition.
2) Davis et al. (1952) attempted to provide a rationale for digital recognition in 1952. The suggested strategy was implemented, and the digital vowel area was shown using the spectacle idea.
3) Olson and Belar (1956) tried to identify ten distinct characters made up of ten monosyllabic words for a single speaker.
4) Fry and Denes attempted to construct a phonetic identifier for analytical concrete, the screen, and similarity in 1959. A hardware-based method surfaced in 1960 when many Japanese laboratories became involved in this area.
5) In 1961, Suzuki and Nakata created hardware for vowel recognition.

Strong synchronisation systems are recommended by Sakoe and Chiba in Japan. Researchers' primary focus in 1970 was on the identification of speech-based discourse in individual words. Line Predictive Coding (LPC) was designed to maximise its visual parameters through the use of voice recognition systems, with an emphasis on low-level coding.

Pruthi et al. (2000) created a word recognition system that relies on word recognition just once. Gupta utilised continuous HMM to see a different name for that Hindi language in 2006. An Arabic speaking system that makes use of HTK and is capable of recognising both isolated and continuous words is Al-Qatab et al. (2010).

- In 2011, R. K. Aggarwal and M. Dave presented the very accurate use of Gaussian reflective mixes for Hindi voice recognition.
- In 2014, Z. Yu and colleagues introduced a voice recognition teaching test that utilised the Hidden Markov Model (HMM) to describe the HMM speech recognition concept and the steps involved in starting a speech recognition programme.

VII. PROBLEM IN TRANSFER RECOMMENDATIONS

The ineffectiveness of the surrounding environment affects how well the voice recognition system works. The degree of active recognition is influenced by a wide range of parameters, including channel variability, speaker/independent reliance, expression level, and environmental influences. However, in order to fill in the gaps using a speech recognition system, it is important to match the changes that occur.

VIII. APPLICATIONS

The usage of digital assistants and speech recognition technology has swiftly expanded from our cell phones to our homes, and its applications in the business, finance, marketing, and healthcare sectors are becoming more and more evident.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

A. AT Work

Speech recognition technology has advanced to the point where it can now execute duties that were previously handled by people, as well as basic tasks that boost efficiency.

The following are a few office duties that digital assistants can or will be able to complete:

1) Look through your computer for reports or papers.
2) Use the data to create a table or graph.
3) Give instructions on the data you wish to have included in a document.
4) Upon request, print documents
5) Launch the video conference call.
6) Set up meetings.
7) Take minutes.
8) Make your trip plans.

B. In The Financial Industry

The financial and banking sectors want voice recognition to make things easier for customers. Voice-activated banking has the potential to significantly cut staff expenses and the requirement for human customer support. In exchange, a customised financial assistant may increase client happiness and loyalty.

Speech recognition's potential to enhance banking by:

1) Without opening your phone, get details about your transactions, balance, and spending patterns.
2) Make the necessary payments.
3) Find out what your transaction history is.

C. In Marketing

Voice-search has the ability to provide advertisers a unique perspective on how to connect with their target audience. Marketers should watch for emerging trends in user data and behaviour given the shift in how consumers will engage with their devices.

1) Data: Marketers will be able to evaluate a new kind of data thanks to speech recognition. Customers' location, age, and other demographic information, such their cultural affiliation, may be inferred from their speech patterns, vocabulary, and accents.
2) Behaviour: Speaking enables lengthier, more conversational searches, whereas typing requires brevity to a certain degree. To keep ahead of these developments, marketers and optimizers might need to concentrate on creating conversational content and long-tail keywords.

Users may become more impatient and reliant on using the internet as their primary information source as a result of this kind of quick search. This may lead to a reduction in the amount of time consumers spend staring at screens. Since there could be a move towards emphasising audio and information-heavy content, marketers should think about what this would entail for content that is mostly visual.

D. In Healthcare

Hands-free, instantaneous access to information can greatly improve patient safety and medical efficiency in a setting when time is of the essence and sterile working conditions are critical.

Advantages consist of:

1) obtaining information from medical records quickly
2) Nurses may receive detailed instructions or be reminded of procedures.
3) Administrative data, such as the number of patients on a floor and the number of units available, can be requested by nurses.
4) Parents may inquire about common sickness symptoms, when to visit the doctor, and how to care for a sick kid from the comfort of their own home.
5) Reduced paperwork,
6) less time spent entering data, and
7) better processes

Regarding voice recognition in healthcare, the biggest worry is what kind of material the digital assistant may access. It is acknowledged that in order for the material to be a practical choice in this sector, it must be provided by and approved by reputable medical organisations.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

E. With The Internet Of Things

With Siri's connectivity to smart thermostats and lights,24 it looks like the days of telling your digital assistant to switch on the kettle are not too far off. The Internet of Things, or IoT, is a significant development that is taking place all around us, not the futuristic potential that it originally seemed to be.

Right now, one of the most well-known uses of voice recognition in the internet of things is in automobiles. By 2020, it's expected that one in five autos will have internet connectivity. With the ultimate goal of reducing driver distractions, the benefits of this might alter the way we interact with and operate our cars. Use of digital assistants in automobiles:

1) Keep your hands free while listening to communications.
2) Manage the radio.
3) Help with direction and navigating
4) Comply with spoken instructions

F. In Language Learning

The potential of voice recognition technology to break down linguistic and cultural barriers in social and professional contexts is among its most revolutionary uses from a human standpoint.

A world free of language barriers creates many opportunities for cooperation across different nations and cultures, which may hasten invention due to the greater variety.

G. Performance Evaluation

The accuracy and speed of the speech recognition technology are typically used to gauge its efficacy. Word error rate is used to measure programme accuracy (WER). As a result, when comparing word error rates (WER), performance is measured in terms of word recognition rate (WRR). The quantity of artefacts, substitutions, and deletions made when viewing the speech can be used to divide name mistakes. These two Word Error Rate (WER) = I + S + D / N is a statistical metric that may be used to assess performance.

The numbers 'I', 'S', 'D', and 'N' represent the input, input, subtraction, and number of words with words, respectively. Word Recognition Rate (WRR) is computed using the formula Real Time Factor (RTF) = T / D, where WER is the speed and the actual time limit is provided. when D is a duration and T is a time series.

IX. CONCLUSION

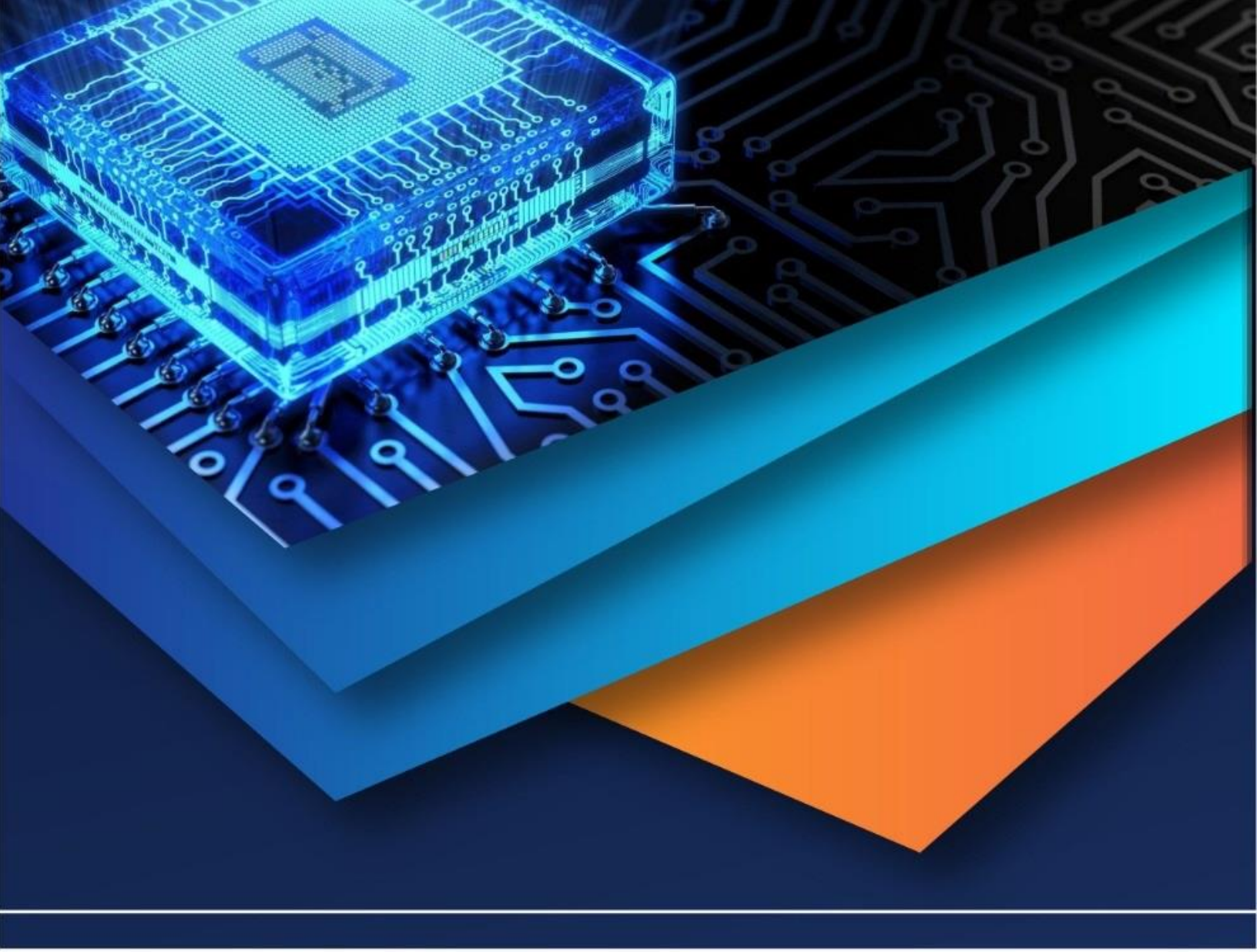
AI has come a long way, and there is still a lot of research being done in this area. The world will become increasingly artificial, and it is up to us as humans to maximise the benefits of AI while reducing the threat to our survival. This study examines their recent development and discusses the fundamentals. This research compares several approaches to developing a voice recognition system using a modified feature translation procedure and speech recognition language system. Voice recognition is a computer study of human speech that primarily aims to translate words and phrases and consistently identify the speaker based on the unique characteristics of their speech waves. Using the presenter's voice is made feasible by this procedure, which also makes personality verification simple. It gives users access to control over a number of services, including safety management, e-commerce, window speak recognition, m-commerce, automation, and home automation. A review of many word recognition speakers and systems is given in this study. Speech is an extremely user-friendly interface as it is a fundamental method of human contact. Despite the industry's increased access to change apps and applications, a few factors still impact the voice recognition system's accuracy and efficacy. The most diversified speech relates to speech volume, speech type, speech channel, and speech context. Certain aspects of the voice signal affect how strong the speech system is. The voice recognition system must be designed with local languages in mind if it is to become more powerful. In the subject of voice recognition, multilingualism is a ground-breaking new area. Many advancements and studies have been made in the field of foreign languages, but it is crucial to employ this technology in the native tongues of indigenous peoples in order to increase its potency and utility. This essay provides a range of speech and speaker detection software.

REFERENCES

[1] International Journal of Artificial Intelligence in Education, 10(2), 130-150. Beck, J.,
[2] Svenmarck, P., Luotsinen, L., Nilsson, M., & Schubert, J. “Applications of artificial intelligence & associated technologies. Science” [ETEBMS-2016] (2018, May).

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

[3] Neuillysur-Seine France. Tao, B., Díaz, V., & Guerra, In Proceedings of the NATO Big Data and Artificial Intelligence for Military Decision-Making Specialists' Meeting (pp. 1-16), Y. (2019).
[4] Artificial Intelligence and Education, Challenges and Disadvantages for the Teacher. Arctic Journal, 72(12), 30-50.
[5] Introduction: Artificial Intelligence for Fashion Industry in the Big Data Era.
In Artificial intelligence for fashion industry in the big data era (pp. 1-6).
Springer, Singapore.
[6] Prerana Das, Kakali Acharjee, Pranab Das and Vijay Prasad "VOICE RECOGNITION SYSTEM: SPEECH-TO-TEXT" JAFS|ISSN 2395-5554 (Print)|ISSN 2395-5562 (Online)|Vol 1(2) |November 2015
[7] M.A.Anusuya and S.K.Katti "Speech Recognition by Machine: A Review" (IJCSIS) International Journal of Computer Science and Information Security, Vol. 6, No. 3, 2009
[8] Nisha Voice Recognition Technique: A Review International Journal for Research in Applied Science & Engineering Technology (IJRASET) ©IJRASET: All Rights are Reserved 262 Volume 5 Issue V, May 2017 IC Value: 45.98 ISSN: 2321-9653
[9] Ashok Kumar, Vikas Mittal "Speech Recognition: A Complete Perspective International Journal of Recent Technology and Engineering (IJRTE)" ISSN: 2277-3878, Volume-7 Issue-6C, April 2019Key Words: Speech recognition, modelling, speech processing, training and assessment., J. B. (2019).



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)