Conversational AI: A Survey

Yash Bhagia[1], Syed Mohammad Abbas[2], Shalendra Kumar[3], Shwetank Maheshwari[4]

[1,2,3,4]*Zakir Hussain College of Engineering and Technology, Aligarh Muslim University, Aligarh (202001) -U.P., INDIA*

---***---

Abstract - *Chatbots, or virtual assistants that can communicate with humans using natural language processing, have grown in popularity in recent years. These computer programs can be used for a variety of things, including finishing duties and offering amusement and business guidance. We start by performing a bibliometric study to determine the key papers and researchers in the field. Then, to spot themes and patterns, we summarize various study articles and present studies in accordance with predetermined criteria. A chronological, thematic, and methodological summary of the related study is also provided. Researchers are also investigating the creation of chatbots that can produce interesting and coherent talks, employing deep reinforcement learning techniques to encourage sequences that exhibit informative, coherent, and simple-to-follow conversational qualities. The way we communicate with machines could be completely changed by these developments in chatbot technology, which also has the potential to completely impact a variety of sectors. An overview of the state-of-the-art in chatbot technology will be provided in this paper, along with suggestions for prospective future research areas.*

***Key Words*: Natural language processing, deep learning, neural networks, chatbots, Conversational Agent, Artificial intelligence.**

1.INTRODUCTION

A subfield of artificial intelligence called conversational AI is concerned with speech-based or text-based AI systems that can replicate and automate verbal interactions with people. Conversational AI is an exciting and rapidly evolving field of artificial intelligence that seeks to develop systems that can engage in natural language conversations with humans. The technology leverages advanced algorithms such as natural language processing, machine learning, and deep learning to understand the meaning of human language and respond in a way that is accurate, contextually relevant, and personalized.

One of the most widely used examples of conversational AI is virtual assistants such as Siri, Alexa, and Google Assistant. These systems enable users to interact with their devices using natural language and voice commands, and the AI technology behind them interprets and processes the user's requests to provide relevant information or complete tasks as needed.

Another example of conversational AI is chatbots, which are computer programs designed to simulate human conversation through text-based interfaces. Chatbots are commonly used in customer service applications, where they can provide support and assistance to customers with basic questions or problems without the need for human intervention.

Conversational AI is also finding innovative applications in various domains such as healthcare, education, and finance. For instance, AI-powered chatbots have been used to provide mental health support to patients, assist with language learning, and provide financial advice to customers.

In healthcare, conversational AI can assist with medical diagnosis and decision-making by analyzing and interpreting clinical data and electronic health records. In education, AI-powered chatbots can provide personalized learning experiences and instant feedback to students, helping to improve their understanding of complex concepts. In finance, conversational AI can assist customers with banking transactions and financial planning, providing personalized recommendations based on their financial goals and needs.

Overall, conversational AI has the potential to transform the way we interact with technology and make our lives more convenient, efficient, and personalized. As the technology continues to evolve, we can expect to see more innovative applications in various domains, making conversational AI an exciting area of research and development.[1]

2. Four Components of Conversational A.I. [2]

2.1. Machine Learning

A subset of artificial intelligence known as machine learning (ML) uses a range of statistical models and algorithms to find patterns and predict future outcomes. Conversational AI needs machine learning to function. It helps the system to improve its understanding of and responses to human language by allowing it to continuously learn from the data it collects. Among other ML subtypes, conversational AI commonly uses supervised learning, unsupervised learning, deep learning, and neural networks.

2.2. Natural Language Processing (NLP)

To create an acceptable answer, natural language processing (NLP) includes converting unstructured input into a machine-readable format. The algorithms that make up conversational AI are always being improved because to the continuous feedback loop that these NLP techniques engage in with machine learning. NLP is crucial to conversational AI since it gives the system the ability to understand human input and generate pertinent responses. To understand human language, there are four fundamental steps to take: Input generation: The process of creating new input for a conversational AI system is known as input generation. Through a website or an app, users can submit text, speech, or both types of input.

Input analysis: This technique identifies the intent and meaning of user input. Conversational AI will employ natural language understanding (NLU) to interpret the substance of the input and determine its intent if it is text-based. To interpret the message upon receiving voice input, it will combine automatic speech recognition (ASR) with natural language understanding (NLU).

Dialogue management: This technique regulates the pace of the conversation by deciding when to clarify or pause. NLG, or natural language generation, creates this function. Conversational AI tracks interactions, determines what information has been acquired, and is still required using dialogue management.

Reinforcement learning: By rewarding or penalizing behaviors, reinforcement learning teaches a system to make decisions on its own. The system is given a goal to accomplish and uses a range of tactics to do so. Depending on how successfully a decision helped it fulfil the task, it receives a score for that decision.

2.3. Data Mining

From vast volumes of data, relevant information is retrieved through data mining. In conversational AI, data mining is utilized to extract patterns and insights from conversational data that programmers can use to improve the operation of the system. Data mining is a technique for discovering undiscovered attributes as opposed to machine learning, which concentrates on making predictions based on recent data despite sharing many characteristics with it.

2.4. Automatic Speech Recognition (A.S.R.)

AI for voice-based conversations includes automatic speech recognition. ASR gives the computer the ability to comprehend human voice inputs, remove background noise, translate spoken words into text, and replicate human responses. Natural language conversations and directed dialogue are the two categories of ASR software.

ASR's streamlined directed dialogue feature may respond to simple yes/no inquiries. ASR interactions that imitate real human conversations are better and more complicated versions of natural language conversations.

Here's a possible paragraph summarizing the sections:

This review paper is organized into nine sections. In the first section, the introduction, we provide an overview of the paper and the significance of conversational artificial intelligence systems. Section II discusses the four components of conversational AI,

including natural language processing, speech recognition, machine learning, and dialogue management. Section III provides an overview of conversational AI systems, including their evolution, current applications, and future potential. Section IV focuses on the system architecture of conversational AI systems, including the different layers and components involved in the process. In section V, we explore the research techniques used to study conversational AI, including data collection, annotation, and evaluation methods. In section VI, we conduct a bibliographic analysis of the literature on conversational AI, identifying key trends and research gaps. Section VII presents the results of our analysis and discussion, highlighting the major findings and implications for future research. Section VIII presents the conclusion, summarizing the main points of the review and providing recommendations for future research in the field of conversational AI.

3. Related Work:

3.1. OVERVIEW OF CONVERSATIONAL ARTIFICIAL INTELLIGENCE SYSTEMS

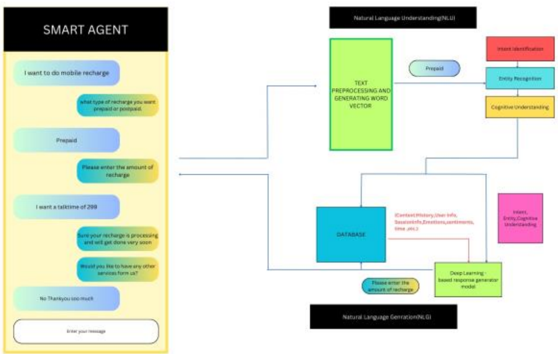
The system architecture of a typical speech-based Conversational AI system as well as a brief history of Conversational AI systems are covered in this section's abridged overview of Conversational AI systems. The introduction of crucial Conversational AI system ideas in this part is crucial. A general comprehension of these ideas is also necessary to comprehend the sections that follow in this essay. As a result, later parts would refer to the terminologies, methods, and ideas presented here.

3.2. History and categorization of Conversational AI systems

Recent years have seen a rapid evolution of conversational AI, giving rise to a variety of terminology to define devices like chatbots, personal digital assistants, and voice user interfaces. Communication methods (voice, text, or a combination), goals (task-oriented or non-task- oriented), initiation and turn (who initiates the conversation), strategies (rule-based, statistical data- driven, end-to-end, or hybrid), and modality (text, speech, or a combination) can all be used to classify these systems.[6] Deep neural networks, statistical methods, and rule-based approaches have all been used in the development of conversational AI systems.[5].

3.3 Rule Based Approach

An early method of system design where decision- making is predefined and manually programmed by designers is the rule-based approach to dialogue management. Although designers may believe they have total control over the system, this strategy has several disadvantages. It is difficult to scale, time- consuming to develop, expensive to maintain, and it might not always produce the best results. The system might also have trouble when users deviate from the planned path, or in edge circumstances.[5]. The adoption of a statistical data-based strategy for dialogue management is prompted by these limitations.



A technique for implementing rule-based AI chatbots was proposed by Nithuna S. and Laseena C.A. The techniques they briefly discussed in their paper include:

- Creating a set of rules to manage user inquiries and produce pertinent responses
- Using regular expressions to check input from the user against established patterns
- Using decision trees to deal with difficult decision- making situations.[8]

The Chatbot Management Process methodology was presented by Santos and Silva as a standardized and iterative way for managing content that can help rule- based AI chatbots develop and evolve. For rule-based AI chatbots, it is essential to establish the chatbot's goals and target market during the methodology's "manage" phase. The "build" phase, which includes human oversight and quality control, focuses on developing and implementing the chatbot is content, including its rules and knowledge base. Through user interactions and feedback, the "analyze" phase assesses the chatbot's performance, allowing for ongoing rule and knowledge base improvement. For rule-based AI, the technique offers an organized approach to chatbot administration. [9]

Mlouk and Jiang provides a methodology for incorporating knowledge graphs into rule-based chatbots, which can improve their ability to understand user queries and generate accurate responses. By leveraging linked data and machine learning techniques, this approach can help rule-based chatbots become more effective in handling complex user queries and provide better user experiences. [10].

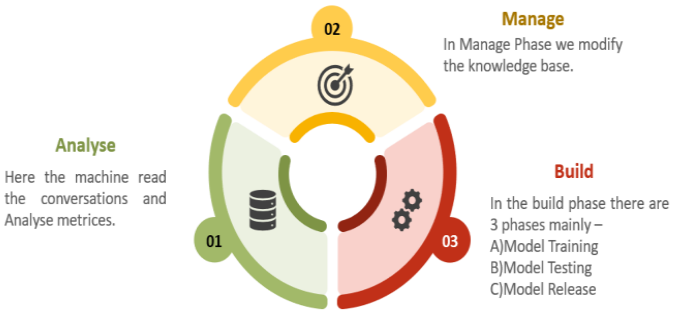
3.4. Statistical data-driven Approach

When creating dialogue management systems, the statistical data-driven approach is employed as an alternative to the rule-based approach. It uses a rule- based approach to probabilistically represent the system's constituent parts and includes a conversation policy, dialogue state tracking, and dialogue control. But this method has significant difficulties, including as scalability, the "black box" issue, and the need for a lot of data to optimize it.

1) 1) Santos and Silva proposed methodology can be seen as a way to incorporate data-driven insights into the development and management of chatbots. By analyzing user interactions and evaluating the chatbot's performance, the methodology allows for the incorporation of statistical data and machine learning techniques to improve the chatbot's performance over time. For example, the authors report While maintaining a high level of confidence in its responses and keeping the user satisfaction data gathered during conversations stable, the methodology's application to Evatalk's chatbot produced positive results, including a decrease in the chatbot's human hand-off rate and an increase in the chatbot's knowledge base examples [9].

2) 2) Yu Wu, Wei Wu, Zhou Jun Li, and Ming Zhou r proposes a method for improving response selection in retrieval-based chatbots using a statistical-based approach. The method uses a weakly supervised learning process that leverages both labeled and unlabeled data. The authors utilize using the sequence (Seq2Seq) model as a crude judgement the matching degree of unlabeled pairs. The experimental results on two public datasets show that this approach significantly improves the performance of matching models. [11].

3) 3) Zehao Lin et al. suggest a novel strategy for dialogue systems to deal with the conundrum of deciding whether to respond to a user's enquiry right away or wait for more details. In addition to introducing the Wait-or-Answer challenge, they also suggest Predict-then-Decide (PTD), a predictive strategy that combines a decision model with two prediction models—a user prediction model and an agent prediction model. The trials carried based on three open datasets and two real-world scenarios demonstrate that the PTD technique beats previous models in resolving this Wait-or-Answer challenge. This study offers a cleverer way for dialogue systems to act in circumstances where they are confused whether to wait or respond, which helps to advance statistically driven conversational AI.[12].



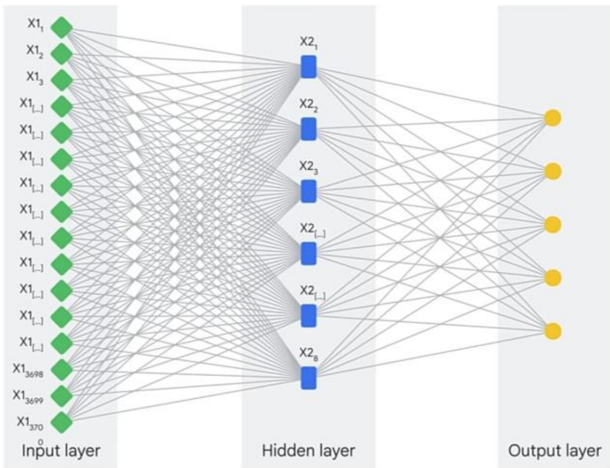
3.5 End-to-End neural approach

Dialogue management system designs modular design can have knock-on consequences, where improving one module has a negative impact on the others. To prevent this, the end-to-end neural approach uses a sequence-to- sequence design that optimizes the entire system; however, this complicates the credit assignment problem, which is the difficulty of identifying underperforming components during evaluation. [7] In spoken dialogue systems, automatic speech recognition and text-to-speech operate independently from the system for input and output, using Deep Neural Networks to map user utterances directly to output.

Vamsi, Rasool proposed that a new method for creating chatbots using a deep neural learning approach. The authors built a neural network with multiple layers to learn and process conversational data. By using deep learning and natural language processing techniques, the proposed approach can improve the accuracy and flexibility of chatbots, allowing them to handle more complex and natural conversations. This research contributes to the advancement of conversational AI by providing a new end-to-end neural approach that can be used to develop more sophisticated and intelligent chatbots for various applications.[13].

Anki, Alhadi as a high accuracy conversational AI chatbot, proposes a bidirectional long short-term memory (BiLSTM) model based on recurrent neural networks (RNN). The pre-processing module, word- embedding layer, BiLSTM layer, attention layer, and output layer are all included in the authors' description of the model's architecture. They also talk about the dataset, which consists of conversational exchanges from a customer service platform, that was utilized to train and test their model. The suggested model outperformed various other baseline models and demonstrated great accuracy in predicting user input replies. The authors contend that more successful chatbots can be created for a range of applications using their BiLSTM-based methodology. [14].

SHEETAL KUSAL1, SHRUTI PATIL, JYOTI CHOUDRIE mentions that neural networks, particularly deep learning models, have been increasingly used for building conversational agents due to their capacity to learn from vast amounts of data and provide more believable and human-like responses. They provide an overview of numerous deep learning architectures, including transformer models, convolutional neural networks, and recurrent neural networks, and how they have been used in conversational agents. [15].



4. SYSTEM ARCHITECTURE OF A CONVERSATIONAL AI SYSTEM

4.1 Automatic speech recognition

The technique known as Automatic Speech Recognition (ASR) transforms audio input from speech to text for further processing. It includes signal and feature extraction, linguistic and auditory models, and hypothesis search. ASR has historically employed N- gram language models, Mel-frequency Cepstral Coefficients (MFCCs), Hidden Markov Models (HMMs), Gaussian Mixture Models (GMMs), and HMMs. Deep neural networks (DNNs), recurrent neural networks (RNNs), and convolutional neural networks (CNNs), among other more recent methods, are being used with better outcomes. A few businesses and institutions, including Google, IBM Watson, and Microsoft Speech API, have developed ASR systems. Word error rate (WER), match error rate (MER), and position independent word error rate (PER) are a few measures that are used to assess ASR performance.

SADEEN ALHARBI, MUNA ALRAZGAN paper briefly explains the working of automatic speech recognition in conversational A.I. system as a user's spoken input is transcribed using ASR and then processed using NLP algorithms to understand the user's intent and generate an appropriate response. Similarly, in a voice assistant application, ASR is used to transcribe the user's spoken commands, which are then used to control the device or perform other actions.[16]

Shinnosuke Isobe, Satoshi Tamura, and Yuuto Gotoh discuss the use of Audio-Visual Speech Recognition (AVSR) in conversational AI in the work. AVSR combines ASR and Visual Speech Recognition (VSR) to increase recognition accuracy in noisy circumstances. In order to determine whether voice data was captured in quiet or noisy environments, the suggested method makes use of a scene classifier based on Parallel- WaveGAN. Only ASR is used in clean surroundings to reduce processing time, whereas multi-angle AVSR is used in noisy conditions to improve recognition accuracy. Using two multi-angle audio-visual databases to test the framework, the study shows that multi-angle AVSR improves recognition accuracy over ASR alone.[17]

In the present investigation, Habib Ibrahim and Asaf Varol explain how automated speech recognition (ASR) identifies and decodes speech signals using an acoustic and modelling technique. An audio signal is transformed into a digital signal, which is then examined to extract properties such the frequency and amplitude of the signal. The speech sounds are then recognized using these traits, and a transcription of the uttered words is produced.

A language model is often used by the ASR system to interpret the words and produce text. The language model uses machine learning and statistical analysis to identify patterns and forecast the likelihood that specific words or phrases will emerge in a particular situation. [18].

4.2 Natural Language Understanding (NLU)

Understanding human language is the primary goal of the Natural Language Understanding (NLU) subfield of Natural Language Processing (NLP). Entity extraction and purpose classification are involved. The technique of determining the user's mood and goal is known as intent categorization. Deep learning techniques like LSTM, CNNs, RNNs, biLSTM, and shallow feedforward networks are increasingly more frequently utilized in place of more conventional methods like Hidden Markov Models (HMMs) and Decision Trees (DTs). The process of identifying entities and categorizing them into predetermined classes is called named entity recognition (NER), sometimes known as entity extraction. While Conditional Random Fields (CRFs) and other conventional methods have been employed, CNNs and biLSTM are currently more frequently used [5].

A thorough assessment and synthesis of the current state of Natural Language Processing (NLP) research in Information Systems (IS) is provided by Dapeng Liu and Yan Li, who provide a roadmap for further NLP research in IS. The authors include 12 commonly studied archetypal NLP tasks, including text classification, information extraction, and sentiment analysis.

The conclusions drawn from this study may be helpful in directing the creation of conversational AI systems. NLP approaches play a significant role in conversational AI systems' ability to comprehend and respond to human input and questions in natural language. Conversational AI developers can take advantage of current research to enhance the performance of their systems by comprehending the status of NLP research and recognizing the archetypal NLP tasks.

4.3. Dialog Management System

The dialogue management system is the key element of a dialogue system that incorporates input from the ASR and NLU and communicates with the knowledge source to decide how the dialogue will go and what steps will be taken after that. It comprises of a decision model or policy model and a context model for discussion state tracking. The context model stores all the dialogue-related data, including any knowledge sources, while the decision model determines the dialogue's progression depending on the data in the context model. The dialogue system's methodology can be rule-based, statistically driven, or neural.

Hayet Brabra and Marcos Baez concentrate on task- oriented conversational systems, which enable users to carry out tasks using the information they disclose during talks. By keeping track of information, comprehending conversation context, resolving ambiguity, managing the conversation flow, interacting with external services/databases, and identifying system actions to achieve the user's ultimate goal, the dialogue management component plays a crucial role in the development of conversational systems.[20]

Takuya Hiraoka and Yuki Yamauchi suggest a technique for creating persuading dialogue systems that engage users in order to accomplish a particular goal while also considering the user's satisfaction. They give an example of a system that guides graduate students in choosing which laboratory to join while also trying to keep the number of students in each laboratory at a steady level. In order to direct users towards issues pertinent to the system purpose, the article suggests techniques for creating knowledge bases and a dialogue management system. The Bayesian network paradigm, on which the persuasive dialogue manager is built, can be used in conversational AI to guide users towards particular activities while still taking their aims and preferences into account.

4.4 Natural Language Generation (NLG)

A dialogue system's NLG component deals with providing the user with the system answer based on the output of the NLU and conversation management system. It is made up of modules for document planning, microplanning, and realization. Responses can be produced using template-based, statistical, neural network, or hybrid methods. The effectiveness of NLG can be assessed using a variety of metrics, including task-based evaluation, grading and human responses are evaluated, and the output is compared to human replies [4].

4.5 Text To Speech Synthesis (TTS)

A conversation system's text-to-speech (TTS) component turns the written output from the NLG into spoken words. The text is analyzed, broken down into phonemes, and then transformed into waveforms as part of this process. Modern TTS methods use neural networks to improve performance over traditional methods that used text and waveform analysis. TTS is assessed using objective, subjective, and behavioral metrics. These include rating synthetic speech, evaluating the effectiveness of user engagement, and gauging the output's intelligibility and comprehensibility.

The usage of a unique Text-to-Speech (TTS) algorithm based on Deep Voice, an attention-based, fully convolutional mechanism, is explored by Pabasara Jayawardhana and Amila Rathnayake. Serenity and fluidity are the most crucial characteristics required from a TTS, with the objective being to provide speech synthesis that is accurate and authentic. The suggested TTS is designed to be used in multilingual synthesizers and performs phonetic-to-acoustic mapping using a neural network-based technique. They also emphasize the value of TTS technology in enhancing web accessibility and user-computer interactions, particularly for people who are visually and audibly handicapped. Given that the majority of listeners are intolerant of unnaturalness, the quality of TTS is essential in keeping a voice that is comparable to that of a human.[22].

Concatenative, formant, articulatory, and hidden Markov model (HMM) synthesis are a few of the text to speech synthesizer techniques covered by Sneha Lukose and Savitha S. Upadhyay. These techniques are used in many applications in the medical and telecommunications domains to produce appropriate sound outputs when text is inputted.

The source-filter paradigm serves as the foundation for the formant synthesis technique, which employs parallel and cascade structures. In general, up to three to five formants are needed to generate understandable speech.

Each formant is represented by a two-pole bandpass resonator with its appropriate formant frequency and bandwidth. While the two-pole resonators in the cascade structure are connected in series, those in the parallel structure are connected in parallel.[23].

In order to make the English grapheme to phoneme dictionary acceptable for Indian English text-to-speech (TTS) systems, Deepshikha Mahanta and Bidisha Sharma suggest a method. A kind of Indian English called Assamese English was given its own set of rules by the authors. By discovering and detecting improbable grapheme-phoneme combinations and correcting them by substituting the most likely vowel phoneme, they improved Indian English's vowel grapheme to vowel phoneme mapping. Both the word level and the sub-word level are refined in this manner.

They implemented the suggested method of dictionary update at the front end of the statistical parametric speech synthesis and unit selection synthesis-based Indian English TTS. After adding the variety-specific information, the subjective assessment revealed a significant improvement in both frameworks.[24].

5. Research Techniques

5.1 Search Strategy: -

With an emphasis on the fields of "Engineering" and "Computer science," the generated search query was employed in Scopus without year limitation, and only documents available in the language is English. The same search query was applied in the IEE Xplore database without limitations (year and publishing themes), and the outputs were sorted using the 'Search inside results' option.

Additionally, the query was limited to the categories "Engineering Civil" and "Construction Building Technology" in the Web of Science Core Collection, without regard to publication years or document types. Because the Boolean connectors in the search option are limited, the search in Science Direct was done sequentially; the 'Engineering' subject area and no year restriction were taken into consideration. Finally, the created query was used to search the ACM digital library.

5.2 Data Extraction and quality assessment: -

In order to gather data for this study, we looked at a wide range of publications (such as books and papers). We examined details such as the publication's creation date, author, subject matter, style, and field of application.

We evaluated each article according to a checklist (similar to a list of criteria) to ensure that they only included high-quality publications in their study. An article was deemed suitable for inclusion in the study if it met at least 70% of the criteria on the checklist.

Quality Assessment Checklist.

No	Checklist
1	Are the aim and objectives clearly stated?
2	Is the reporting logical and coherent?
3	Are the proposed technique well described?
4	Is the employed research methodology suitable for the objectives?
5	Are the methods for data collection adequately described?
6	Do the interpretation and conclusion hinge on the data?
7	Is there an incremental contribution to knowledge?
8	Are the aims and objectives fulfilled?
9	Is the research process well documented?
10	Is the study reproducible?

6. Bibliographic Analysis

6.1. Publication Trend:

The publication trend of conversational A.I. shows a steady increase over the last decade. The number of publications on this topic increased from 1 in 1998 to 1586 in 2023. The year 2023 witnessed the highest number of publications (1586) on this topic. The trend shows that conversational A.I. is a rapidly growing area of research.

6.2. Top Countries:

The United States is the leading country in the publication of research related to conversational A.I., with 1686 publications. It is followed by China (1338), the United Kingdom (374), Germany (336), Canada (275), France (259), Italy (242), Spain (232), Japan (227), India (123).

6.3. Top Authors:

The top author in this field is Xiodan Zhu from Queens, with 26 publications. He is followed by Liu, Bing from the University of Windsor in Canada. with 23 publications, and Dr Shivani Agarwal from Indian Institute of science with 14 publications.

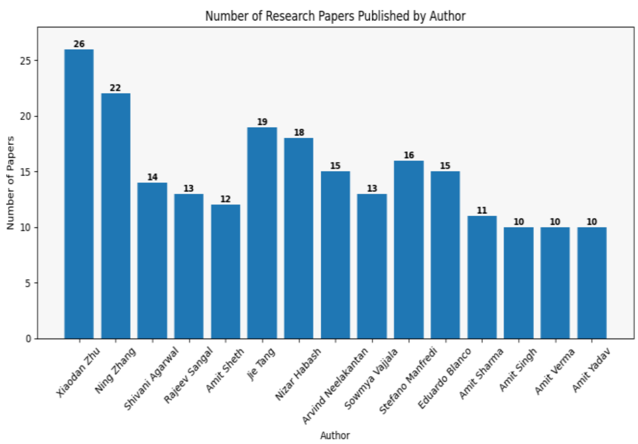
6.4. Top Journals:

With 130 publications, the top journal for conversational A.I. research is IEEE Transactions on Audio, Speech, and Language Processing. Journal of Machine Learning Research (36), and ACM Transactions on Interactive Intelligent Systems (57) follow it.

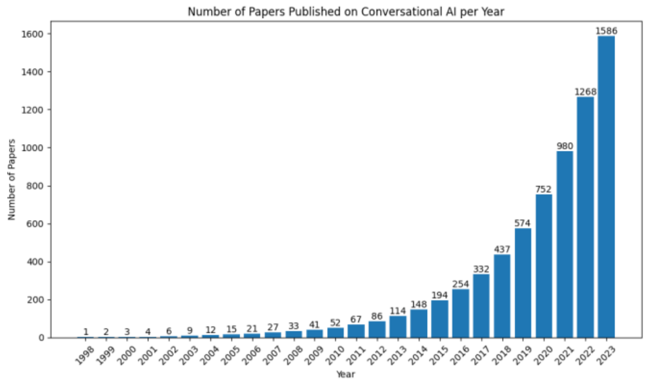
6.5. Top Keywords:

The top keywords in this field are "natural language processing" (1,623 publications), "chatbot" (893 publications), "dialogue system" (702 publications), "machine learning" (658 publications), and "speech recognition" (608 publications) [25].

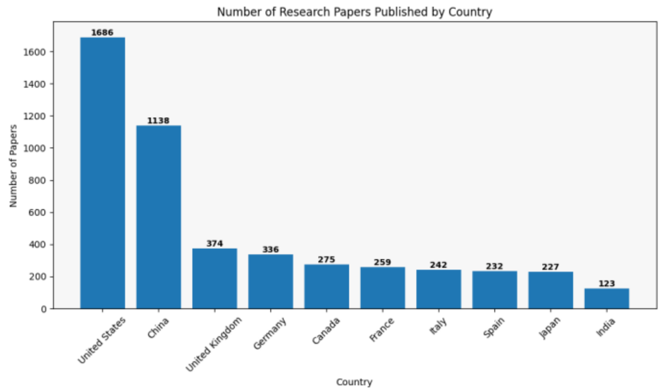
Documents by Author



Documents By Year



Documents by Countries



7. Result

7.1. A critical analysis of developed conversational AI systems: -

This section provides an overview of the conversational AI systems' evolution, as well as its evaluation and breakdown of their components. This section is essential as it provides a broad overview of the elements and techniques used to build conversational AI systems.

Although the components have been broken down into ASR, NLU, Knowledge source, Dialogue management, NLG, and TTS, depending on the strategy and aim of the system, not all Conversational AI systems have employed these in their development. While a modularized system used many components for various purposes, the seq-2-seq just used one component for more than one task.

The ASR's ability to transform voice into text, from which the NLU Unit derives the requests of the users, is crucial. Both syntactic and semantic analysis are used in this procedure. Parsing, also known as syntactic or syntax analysis, is the process of dissecting material to see whether it complies with grammar rules.

On the other side, semantic analysis works with deriving meanings from the words to identify the request from the users, and it frequently makes use of knowledge bases or domain ontologies. Before generating any queries, the NLU uses domain ontology to give the machine a field understanding of the meaning of the word. A domain ontology must be created and maintained, which takes time.

Additionally, because the majority of the systems under consideration are designed to facilitate information retrieval, the DM oversees and controls the procedure for obtaining the appropriate data from the database in response to user requests.

For the instance, some systems used vectorization and the mapping of the search terms to the database. The interaction between more than one component, such as the interaction between the BIM model and dynamo [49,55,57] used for data extraction and manipulation of the BIM model, might be directed on a platform that has been established based on pre-coded scripts. After the activities are finished, the users receive feedback in the form of textual, visual displays, or speech. The answer frequently used a template to provide consumers' feedback.[27][28][29][30][31].

7.2. System verification and assessment: -

All of the research that was evaluated was validated, with the exception of the framework proposed by Sheldon and Dobbs [61]. With the assumption that happy path users will test with anticipated inputs and anticipate anticipated outcomes, the suggested Conversational AI systems are still in the early stages of research and validation in a controlled setting. The validation process may involve testing using a case study, a task, a group of queries, or a series of questions, depending on the system's objective. nearly all of the recommended

Systems were evaluated based on user reviews, performance tests, or contrasts with other current systems. Users are asked a series of questions about how they use the system in order to provide feedback (satisfaction, efficiency, and effectiveness) for the user experience evaluation. On the other side, performance evaluation compares the output of the system to recognized benchmarks or the real world using established criteria. [29][30][31].

7.3.Validation and evaluation of the systems: -

They examined whether each system had been verified (tested to ensure appropriate operation) and evaluated (assessed for performance).

Except for the Sheldon, Dobbs framework, the majority of the systems had their validity checked.

The researchers discovered that the majority of Conversational AI systems were still under development and were being evaluated in a controlled setting. This indicates that the systems were tested using expected inputs (such as inquiries or commands that the developers were aware would likely be given) and responded accordingly.

The researchers searched for information regarding how the systems would function in edge scenarios, but they couldn't uncover any. Edge cases are scenarios in which the system receives unexpected inputs

(such as inquiries or orders that the developers weren't expecting).

The majority of the suggested Conversational AI systems were assessed by the researchers in three different ways: user experience, performance, and comparison with existing systems.

User experience evaluation refers to the researchers' inquiry into the satisfaction, effectiveness, and efficiency of system from those who actually utilized it.

The term "performance evaluation" refers to the researchers' comparison of the system's output to a benchmark or standard. In one study, for instance, the system's performance was assessed based on how well it handled particular tasks (such as calculating writing and reading latency in a blockchain-based network). Other studies assessed the system's performance using metrics like recall, accuracy, and precision.

8. Conclusion

The interaction between people and machines has improved over time thanks to conversational AI. This study conducted a thorough assessment of Conversational AI systems in the construction industry in order to identify the present application areas, the development of the systems, prospects, and issues. A thorough search of Scopus, IEEE Xplore (Institute of Electrical and Electronics Engineers), ACM (Association of Computing Machinery), Web of Science, and Science Direct turned up numerous publications on Conversation AI systems. In the current work, these were retrieved and rigorously examined.

Most Conversational AI systems in the examined studies used a modularized design. Like this, most ASR systems used Google Speech Recognition System. Information extraction and retrieval are the main current uses of conversational AI systems. Current conversational AI systems concentrated on retrieving or extracting data from websites, project schedules, building information models, and other sources. All these systems are task-oriented and intended for use by architects, building managers, building occupants, teachers, and construction workers. The goals of these Conversational AI systems are to enhance user experience, increase productivity by cutting down on contact time, and enhance cognitive learning.

Last but not least, this study has limitations despite its contributions. Data from Scopus and Web Science were used in the study, and they were later verified using databases from IEEE Xplore, ACM, and Science Direct. Thus, the study's ability to cover various datasets might be constrained. Similar to this, the selection criteria and the search queries could both act as restrictions. However, when doing the systematic review, best practice guidelines were used. To the best of the authors' knowledge, this research is significant since it represents the first attempt to provide an overview of conversational AI.[27][28][29][30][31].

REFERENCES

[1]. https://www.ibm.com/in-en/topics/conversational- ai

[2]. https://yellow.ai/conversational-ai/

[3]. M. Campbell, Beyond Conversational Artificial Intelligence, Computer 53 (12) (2020) 121–125

[4]. S.A. Bello, L.O. Oyedele, O.O. Akinade, M. Bilal, J.M. Davila Delgado, L.A. Akanbi, A.O. Ajayi, H.A. (2021) 122.

[5]. M. McTeer, Conversational AI: Dialogue Systems, Conversational Agents, and Chatbots, Synthesis Lectures on Human Language Technologies. 13 (2020)

[6]. Magalhaes, J., A.G. Hauptmann, R.G. Sousa, and

C. Santiago, MuCAI'21: in Proceedings of the 29th ACM International Conference on.2021.

[7]. A.C. Morris, V. Maier, P. Green, From WER and RIL to MER and WIL:

[8]. Nithuna S and Laseena C.A.: Review on implementation Techniques of Chatbot (2020).

[9]. GIOVANNI ALMEIDA SANTOS, GUILHERME GUY DE ANDRADE, GEOVANA RAMOS SOUSA SILVA 1, FRANCISCO CARLOS MOLINA DUARTE, JOÃO PAULO JAVIDI DA COSTA

, RAFAEL TIMÓTEO DE SOUSA, A Conversation driven Approach for Chatbot Management. (2022).

[10] Addi M Mlouk, Lili Jiang: A knowledge-based graph chatbot for Natural language understanding over Linked Data. (2020).

[11] Yu Wu, Wei Wu, Zhoujun Li, and Ming Zhou.: Learning Matching Models with Weak Supervision for Response Selection in Retrieval-based Chatbots (2019).

[12] Zehao Lin, Shaobo Cui, Guodun Li, Xiaoming Kang, Feng Ji, Fenglin Li, Zhongzhou Zhao, Haiqing Chen, and Yin Zhang: Predict-then-Decide: A Predictive Approach for Wait or Answer Task in Dialogue Systems (2021).

[13] G Krishna Vamsi, Akhtar Rasool, Gaurav Hajela: A Deep Neural Network Based Human to Machine Conversation Model. (2020).

[14] Prasnurazaki Anki, Alhadi Bustamam, Herley Shaori Al-Ash, Devvi Sarwinda: High Accuracy Conversational AI Chatbot Using Recurrent Neural Networks Based BiLSTM Model. (2020)

[15] SHEETAL KUSAL1, SHRUTI PATIL, JYOTI CHOUDRIE, KETAN KOTECHA, SASHIKALA MISHRA, AND AJITH ABRAHAM: AI-Based Conversational Agents: A Scoping Review from Technologies to Future Directions (2020).

[16] SADEEN ALHARBI, MUNA ALRAZGAN, ALANOUD ALRASHED, TURKIAYH ALNOMASI, RAGHAD ALMOJEL, RIMAH ALHARBI, SAJA ALHARBI, SAHAR ALTURKI, FATIMAH ALSHEHRI, AND MAHA ALMOJIL.: AutomaticSpeech Recognition: Systematic Literature Review.

[17] Shinnosuke Isobe, Satoshi Tamura, Yuuto Gotoh and Masaki Nose: Efficient Multi-angle Audio-visual Speech Recognition using Parallel WaveGAN based Scene Classifier (2022).

[18] Habib Ibrahim, Asaf Varol: A Study on Automatic Speech Recognition Systems.

[19] Dapeng Liu and Yan Li: A Roadmap for Natural Language Processing Research in Information Systems.

[20] Hayet Brabra, Marcos Baez, Boualem Benatallah, Walid Gaaloul, Sara Bouguelia, Shayan Zamanirad: Dialogue Management in Conversational Systems: A Review of Approaches, Challenges, and Opportunities (2021).

[21] Takuya Hiraoka, Yuki Yamauchi, Graham Neubig, Sakriani Sakti, Tomoki Toda, Satoshi Nakamura: DIALOGUE MANAGEMENT FOR LEADING THE CONVERSATION IN PERSUASIVE DIALOGUE SYSTEMS.

[22] Pabasara Jayawardhana, Amila Rathnayake: An Intelligent Approach of Text-To-Speech Synthesizers for English and Sinhala Languages (2019).

[23] Sneha Lukose, Savitha S. Upadhya: Text to Speech Synthesizer-Formant Synthesis.

[24] Deepshikha Mahanta, Bidisha Sharma, Priyankoo Sarmah, S R Mahadeva Prasanna: Text to Speech Synthesis System in Indian English.

[25] https://www.scopus.com/home.uri

[26] SHEETAL KUSAL1, SHRUTI PATIL, JYOTI CHOUDRIE3, KETAN KOTECH, SASHIKALA MISHRA, AJITH ABRAHAM: AI-based Conversational Agents: A Scoping Review from Technologies to Future Directions.

[27] K. Adel, A. Elhakeem, M. Marzouk, Chatbot for construction firms using scalable blockchain network, Autom. Constr. 141 (2022).

[28] B. Zhong, W. He, Z. Huang, P.E.D. Love, J. Tang, H. Luo, A building regulation question answering system: A deep learning methodology, Adv. Eng. Informatics (2020) 46.

[29] S. Wu, Q. Shen, Y. Deng, J. Cheng, Natural- language-based intelligent retrieval engine for BIM object database, Comput. Ind. 108 (2019) 73–88.

[30] G. Gao, Y.-S. Liu, M. Wang, M. Gu, J.-H. Yong, A query expansion method for retrieving online BIM resources based on Industry Foundation Classes, Autom. Constr. 56 (2015) 14–25.

[31] Zhoui, H., M.O. Wong, H. Ying, and S.H. Lee, A Framework of a Multi-User Voice Driven BIM-Based Navigation System for Fire Emergency Response, in 26th International Workshop on Intelligent Computing in Engineering (EG-ICE 2019). 2019.