iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** IV **Month of publication:** April 2024

DOI: https://doi.org/10.22214/ijraset.2024.60231

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 IV Apr 2024- Available at www.ijraset.com

A Deep Dive into the Technology Driving Modern Chatbots: A Review

Saurav Amolik
Sonopant Dandekar Arts, V S Apte Commerce & M. H. Mehta Science College Palghar

Abstract: ***Chatbots have always been an interesting topic for research based on the technology used to build them. A chatbot is an intelligent system which can bring out conversations with a human using humanly or natural language in real time. Many businesses now use online platforms to reach out to the customer inquiries, while many of them turn to chatbots for improving their customer service and increasing their productivity. Based on various research being done on chatbots, some important issues regarding them have changed over time. Therefore, it is necessary to review some of the principal technologies most liable to build highly independent and self-learning chatbots. This paper introduces a comprehensive discussion about the technologies with a focus on recent advancements. At first, this paper will provide an overview of chatbots and then focus on technologies used to build the chatbots. It will also grab your attention towards the in demand chatbots built in the recent years and how they differ with each other in terms of functionalities.***

Keywords: Chatbots, Natural Language Processing (NLP), Pattern Recognition, Semantic Web, Data Mining, Transformer-Based Models.

I. INTRODUCTION

Human-machine conversation is a technology that integrates different areas to facilitate communication between users and computer using natural language. A chatbot is a related term to machine conversation, a conversational agent that interacts with users using natural language [1]. The interaction is done mostly through voice and text conversations. Modern chatbots are usually online and use generative artificial intelligence systems that are able to converse with the user in natural language and simulate the behaviour of a person as an interlocutor. Such chatbots often use deep learning and natural language processing technologies. As of late 2022, the field has gained a lot of attention, with variants such as OpenAI's ChatGPT, Microsoft's Copilot, and Google's Gemini gaining popularity [2]. Among the language-intelligence technologies, chatbots are currently receiving valuable attention in academic and practical areas [3]. A major area where chatbots have long been used is in customer service and support. Users can get the information they need by sending a message to the chatbot service without making a request or visiting a website. Therefore, many companies are offering various services to customers by customizing chatbots to match their marketing and advertising strategies [3].

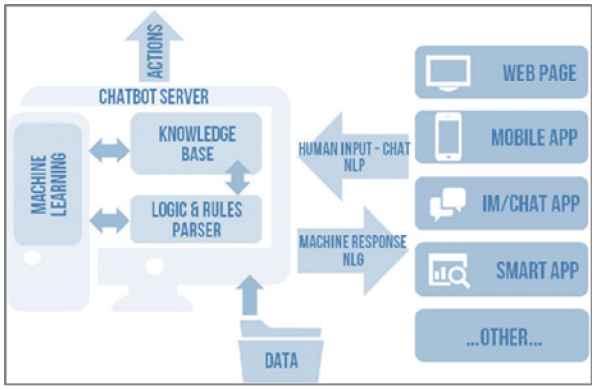


Fig:1 The structure of chatbot (Image Courtesy: Martin Hasal[1] Jana Nowaková[2] Khalifa Ahmed Saghair[2] Hussam Abdulla[2] Václav Snášel[2] Lidia Ogiela[3] [4])

One of the basic challenges of chatbot is to train machines to communicate with humans using natural language. The basic concept of chatbot was proposed by Alan Tune in 1950, and in 1966, the chatbot called as ELIZA mimicked human dialogue. However, this gave rise to lot of questions and limitations. Among the most notable early chatbots are ELIZA (1966) and PARRY (1972) [2].

©IJRASET: All Rights are Reserved

Later in 1995, ELIZA tried to construct a dialogue pattern using the Artificial Markup language (AIML). but still there were number of limitations in this technology. More similar attempts have been made till the late 90s and early 2000 with the primary aim to build a computer program that could successfully mimic human conversations. Since the past 8-10 years, modern chatbots are highly dependent on the technologies such as Natural Language Processing (NLP), Deep Learning, Pattern Matching, Semantic web and many more. The chatbot business is growing due to the popularity of messaging apps among the new millennial generation and advancements in AI-related technologies. Currently, WhatsApp and Facebook Messenger have the largest global market share, while WeChat and LINE have similar shares in China and Japan [3]. In 2016, Facebook Messenger developers created 100,000 bots for Messenger in the first year of the Messenger Platform. The initial Messenger Platform, announced onstage at F8 conference a year ago, was created to allow businesses and third-party developers to make their own bots. It includes chat extensions, tabs dedicated to the exploration of bots and local businesses, and custom integrations with apps like Delivery.com [5].

Modern chatbot design can generally be classified into several key elements, which are knowledge, response generation, text processing, machine learning (ML) model that are usually based on neural networks [26]. Chatbots seem to provide users with quick and convenient support to answer their questions. A common motivation for chatbot users is productivity, while other motives include entertainment, social factors and exposure to new things [27]. However to balance all this motivations, a chatbot must be built to function as a tool, a toy, and a friend [27][28]. The main task of a chatbot is to generate an appropriate response based on natural human input. There are several ways to create a response, which defines the modeling mechanism of the chatbot, which is shown in Fig:2 [29]. One is the rule-based method, wherein these rules are usually set by developers and determine how the chatbot interacts with users based on the keywords or patterns they enter. This type of chatbots can be effective in scenarios where communication is relatively simple and predictable, such as providing basic customer support or answering common questions. The other one is neural-network-based approach which relies on deep learning. Neural network-based chatbots, are also known as AI-based chatbots that use artificial neural networks to understand and generate human-like responses in conversations. Neural network-based approaches can be broadly classified as retrieval based and generative. Retrieval based methods generate an answer by computing the most relevant answer, rather than based on a score-function method, such as calculating conditional probabilities [29][30] implemented by a neural network or evaluating the relationship between the context and candidate answers in an enhanced version [29][31]. On the contrary, Generative method generates one word at a time corresponding to the given input after probabilities have been computed over the whole vocabulary [29][32].

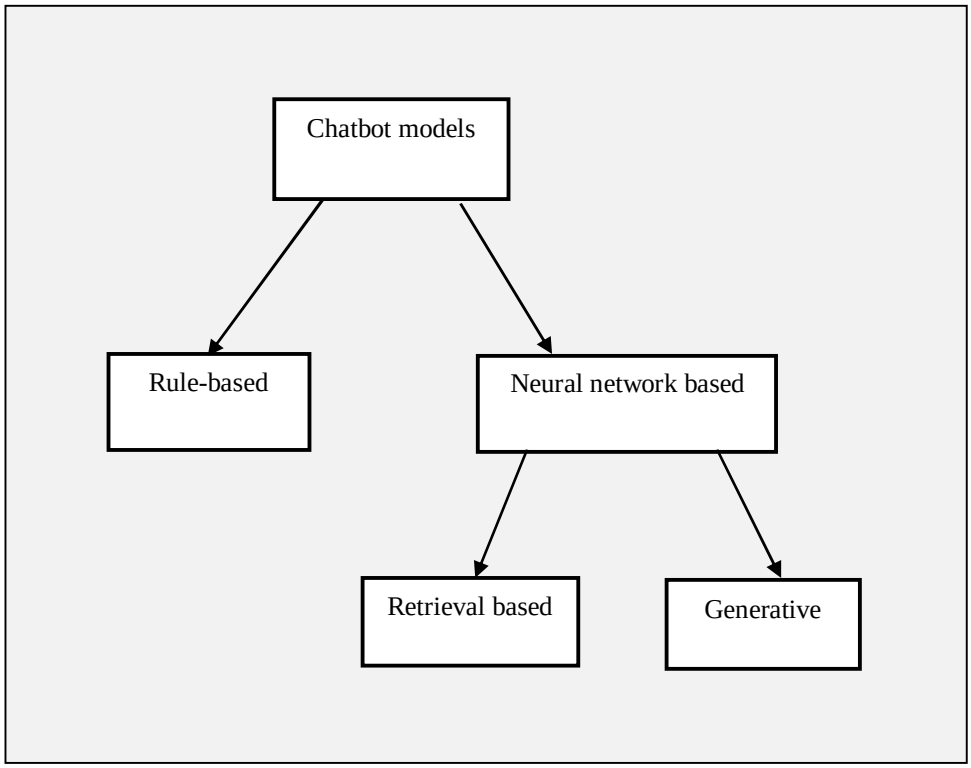


Fig:2 Classification of chatbot models

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

II. TECHNOLOGIES AND THEIR ADVANCEMENTS

A. Natural Language Processing (NLP)

Natural language processing (NLP) is a branch of artificial intelligence (AI) that focuses on enabling computers to understand, interpret and reproduce human language in a natural and meaningful way. It involves the development of algorithms and models that enable computers to understand various aspects of human language, including symbols, grammar, and context. This includes tasks such as tokenization, part-of-speech tagging, parsing, natural entity recognition and sentiment analysis. The goal of NLP systems is to understand the meaning of human language. This involves processing the text at a deeper level to make sense, extract relevant information and respond appropriately. Natural language Understanding (NLU) enables applications such as virtual assistants, chatbots, and machine translation systems to understand and respond to user queries [3]. In addition to understanding human language, NLP also includes generating similar sounding text, which is also called as Natural Language Generation (NLG). In 2023, the StartUs insights published an article that overviewed the top global natural language processing trends. The top 9 trends included Virtual Assistants, Sentiment Analysis, Multilingual Language Models, Named Entity Recognition, Language Transformers, Transfer learning, Text Summarization, and Semantic Search Reinforcement Learning. The article's aim was to handle the issues such as analyzing vast user data to generate insights for the rapidly growing social media and also the chatbots were unable to fully replicate humans due to the lack of understanding of interpretations and context [6].

B. Pattern Recognition

Pattern recognition are classification techniques that allow suitable description and identification of data sets of interest to the user based on a particular pattern description or expression. The main characteristics of most pattern recognition techniques are the notions of generality and specificity [7]. Pattern recognitions is observed as one of the most significant processes of machine learning [8]. Pattern recognition processes can be divided into supervised learning, unsupervised learning, and semi-supervised learning processes. It is actively used in computer diagnostics, speech recognition, identification of health interpretations, and automated recognition of photographs, images, faces, handwritten images, and other analysis. Chatbots based on pattern matching are usually called as rule-based chatbots that rely on predefined patterns or rules to understand respond to user inputs. These rules are created by human developers. When a user provides an input to the chatbot system, the system scans the user's input for matching keywords or phrases. If it identifies a match, the chatbot selects a predefined response associated with that pattern or rule. Deep learning techniques such as convolutional neural networks (CNN), recurrent neural networks (RNN), and transformer-based architectures such as BERT and GPT, have shown significant success in pattern recognition tasks [9]. These models have expertise in tasks such as image recognition, speech recognition and natural language understanding, thus achieves high performance in various benchmarks [9]. Following example clearly defines how pattern recognition works while training the rule based chatbots [10].

```
#PATTERN MATCHING
<category>
   <pattern>What is the capital of \*</pattern>
   <template>
      <srai>The capital of <star/></srai>
   </template>
</category>
-------------------------------------------------
#PRE_STORED PATTERNS
<category>
   <pattern>The capital of the United States of America?</pattern>
   <template>The capital of the United States of America is Washington, D.C.</template>
</category>
<category>
   <pattern>The capital of India?</pattern>
   <template>The capital of India is New Delhi.</template>
</category>
```

©IJRASET: All Rights are Reserved

C. Semantic Web

The Semantic Web, also known as Web 3.0, is an extension of the World Wide Web through standards established by the World Wide Web Consortium (W3C). The goal of the Semantic Web is to make Internet data machine-readable. To enable the encrypting of semantics with the data, technologies such as Resource Description Framework (RDF) and Web Ontology Language (OWL) are used [11]. These technologies mostly represent metadata. The Semantic Web can play an principal role in improving the capabilities of chatbots by helping them better understand and process information from various web sources. By understanding data semantics, chatbots can aggregate information from multiple sources and provide users with more comprehensive and accurate responses. Chatbots can be enabled to use SPARQL queries to search RDF datasets and provide users with more accurate and contextual answers. Ontologies can help chatbots understand the context of user queries and conversations by recognizing relationships between concepts and entities. This may allow chatbots to provide accurate and personalized responses based on the user's interest and previous interactions. Semantic Web standards support interoperability between different systems and data sources by providing a common framework for specifying and exchanging data. Chatbots can use this connection to access and integrate data from multiple sources, including databases, APIs, and web services, to better adapt to different environments [13]. The trends of semantic technology for 2024 will be based on several technological improvements. The introduction of ChatGPT has created various primary and secondary effects on semantic technology and the industries using it [12].

D. Data Mining

Data mining is the process of extracting useful information and patterns from big data. This is also called knowledge discovery, knowledge mining from data, knowledge extraction, or data/pattern analysis [14]. It is basically a process that searches logically through large amount of data to find useful data. The aim of this technique is to find patterns that were previously unknown. Once data is collected, it often needs to be cleaned and transformed to make it ready for analysis. This includes handling missing values, removing duplicates, comparing, or multiplying numeric values, and encoding categorical variables. Data mining can be considered as a valuable tool for enhancing chatbots.

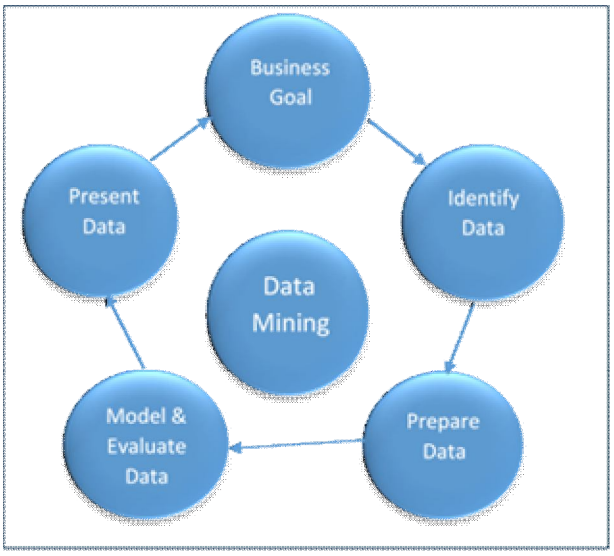


Fig:4 Data Mining Cycle

Data mining uses algorithms and many other techniques to transform large data sets into useful output. The most popular data mining techniques are association rules, classification, clustering, decision trees, K-nearest neighbour, neural networks, and predictive analysis. By mining user data, chatbots can tailor interactions based on a user's individual preferences, behaviour, and past interactions. This may include recommending relevant content, products or services based on the user's history and preferences. For example, if a chatbot detects that a user frequently asks about a certain topic, it can tailor its response to provide more relevant information. Its scope has expanded by hybridizing various data mining algorithms used in fintech and cryptocurrencies, blockchain, data science, sentiment analysis and recommendation systems. In addition, data mining offers advantages in many practical fields such as health data analysis and mining, biology, information security, smart cities, and privacy preservation of smart networks [15].

E. Transformer-Based Models

Transformer-based models include OpenAI's GPT (Generative Pretrained Transformer) series and Google's BERT (Bidirectional Encoder Representations from Transformers). These technologies have completely revolutionized the functions of chatbots in the entire world market. Generative Pre-trained Transformers are neural network models that use the transformer architecture and are a key advancement in artificial intelligence (AI) powered applications such as ChatGPT. GPT-3 (Generative Pre-trained Transformer 3) is a large language model released by OpenAI in 2020. Like its predecessor, GPT-2, which was just a decoder [16]. GPT models allow applications to create human-like text and content, and answer questions in a conversational manner. Various industries are using GPT models and generative AI for Q&A bots, text summarization, content generation, and search [17].

Google's BERT (Bidirectional Encoders for Transformers) is a language model based on the transformer architecture, which is a significant and dramatic improvement over previous state-of-the-art models. It was introduced in October 2018 by Google researchers [18]. BERT was originally implemented in English in two model sizes: (1) BERTBASE: 12 encoders with 12 two-way self-monitoring heads for a total of 110 million parameters, and (2) BERTLARGE: 24 encoders with 16 two-way self-monitoring heads for a total of 340 million parameters. Traditional language models, such as those based on recurrent neural networks (RNNs) or convolutional neural networks (CNNs), process words sequentially or with context windows of fixed size, respectively. However, BERT uses a transformer architecture that allows the entire input phrase to be examined at once. BERT is pre-trained on large text datasets using two unsupervised tasks: masked language model (MLM) and next sentence prediction (NSP). The MLM task involves masking some words in an input sentence and training a model to predict those masked words based on the context of the rest of the sentence. The NSP predicts whether one sentence follows another in each body of text. The below Table (1) generalizes the differences between Google's BERT and OpenAI's GPT [20].

	BERT	**GPT**
Architecture	BERT is designed for bidirectional representation learning. It uses the purpose of a masked language model where it predicts missing words of a sentence based on both left and right context.	GPT, on the other hand, is designed for generative language modelling. It predicts the next word of the sentence in the previous context using a one-way autoregressive approach.
Pre-training Objectives	BERT is pre-trained using a masked language model target and next sentence prediction. It focuses on capturing two-way context and understanding the relationships between words in a sentence.	GPT is pre-trained to predict the next word in a sentence, which prompts the model to learn a coherent representation of the language and generate contextually meaningful sequences.
Context Understanding	BERT is effective for tasks that require a deep understanding of sentence context and relationships, such as text classification, called entity recognition, and question answering.	GPT is strong in creating coherent and contextual text. It is often used in creative tasks, dialog systems, and tasks that require the generation of natural language sequences.
Task types and Use Cases	Often used for tasks such as text classification, named entity detection, sentiment analysis, and question-answering.	Applied in tasks such as text generation, dialogue systems, summarizing and creative writing.
Fine-tuning vs Few-Shot Learning	BERT is often fine-tuned for specific subsequent tasks, with information tagged to adapt pre-trained representations to the task at hand.	GPT is designed to be learned a number of times, where it can be generalized to new tasks with minimal task-specific training data.

Table (1): BERT VS GPT

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

III. CHATBOT STATISTICS IN 2024

According to a Demandsage article published by Rohit Shewale, this figure shows the chatbot market revenue growth from 2018 and forecast growth till 2027 [25]. The chatbot sector is estimated to be worth $179.9 million by 2024. Additionally, by 2027, its revenue is expected to be $454.8 million.

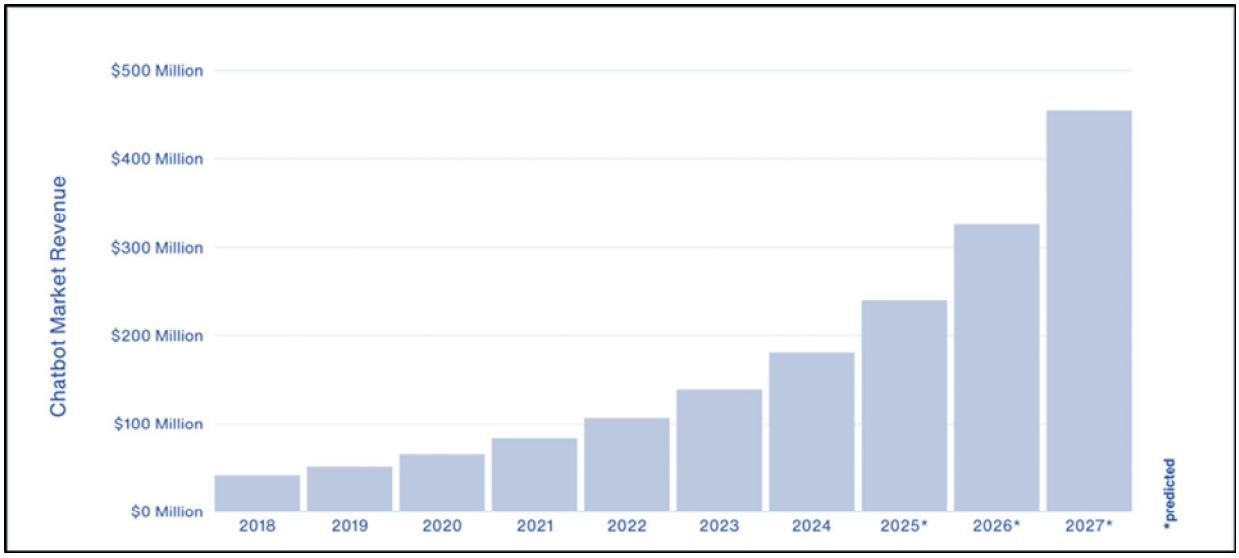


Fig:5 Chatbot Market Revenue between years 2018-2027 [25]

IV. BEST AI CHATBOTS IN RECENT YEARS

After exploring the best and most useful technologies for creating effective chatbots, the research moves on to the most popular AI chatbots of the last 5-6 years. A recent article titled "Best AI Chatbots: ChatGPT Isn't the Only One Worth Trying" written by ZDNET editor Sabrina Ortiz featured some of the most popular, reliable, and accurate AI chatbots. Microsoft Copilot, ChatGPT, Anthrophic's Claude, Jasper, Gemini are some of the in-demand AI Chatbots.

A. Microsoft's Copilot

In February 2023, Microsoft launched its AI-powered Copilot (previously named Bing) which uses OpenAI's Large Language Model (LLM). Copilot operates on GPT-4 Turbo, the advanced technology of OpenAI. As of May 2023, Copilot went from limited preview to open preview, meaning everyone can now use it for free. After using both ChatGPT and Copilot, the study concludes that Copilot uses the power of AI specifically to help with coding, while ChatGPT offers a wider range of general-purpose language processing capabilities.

B. ChatGPT

ChatGPT is the fastest growing AI chatbot all around the world with nearly 180.5 million users. It was released in November 2022, and its massive success is the biggest motivation for other chatbots to come into existence. ChatGPT has earned the name "Original Chatbot". It is trained in such a manner that it has answer for almost every query or question asked by a user. It can literally write for you letters, emails, essays, grocery lists, poems and much more. In addition to these positive features of ChatGPT, the analysis of the study found an issue that ChatGPT is trained on a large data set, which can lead to biased responses. Also, you cannot expect to get the same relevant response, every time you ask the same query again and again.

C. Anthrophic's Claude

Anthrophic launched its first AI chatbot, Claude, in February 2023. It is similar like other chatbots in answering anything you need help with, such as coding, writing, research, and many more. One of the biggest advantages is we can upload any file we wish to read, analyze and summarize. It can transcribe and analyze static images such as diagrams, handwritten notes and photographs. ClaudeAI's biggest limitation is simply the amount of information it contains. Because ClaudeAI do not browse the web or build data dynamically, its knowledge is fixed during training.

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

That means that it can fall short answering some tricky questions. Its knowledge breaks at the beginning of 2022, so very recent current events also create gaps [22]. "ClaudeAI Pros And Cons [2024]" January 12, 2024 by Claude Ai deeply specifies key advantages and limitations of Claude AI [22].

D. Jasper

Jasper AI is usually best for business and marketers. It does the work of checking plagiarism in documents, condenses texts and creates paragraphs and product descriptions [21]. you can enter a prompt of what you want to type in Jasper and it will type it for you, just like ChatGPT. But the major difference with Jasper is that it has a wide range of tools to make a better copy [21]. It can be a great alternative for ChatGPT, but the most important thing to remember is that Jasper AI is a tool that enhances your skills, not replaces them. The content it produces still needs checking, editing, and proofreading [23].

E. Gemini

Gemini is Google's conversational AI chatbot is developed in March 2023. It was formerly known as Bard. Gemini works like Copilot, pulling its responses from the web, providing footnotes, and even creating images within its chatbot. It combines world knowledge with information extracted from images and videos [21][24]. It has the ability to answer questions and extract information from various contents like infographics, charts, figures, tables, and web pages. Recently Gemini faced a downside in February 2024, when social media users and conservative commentators reported that it was producing historically inaccurate images of historical figures as people of color [24]

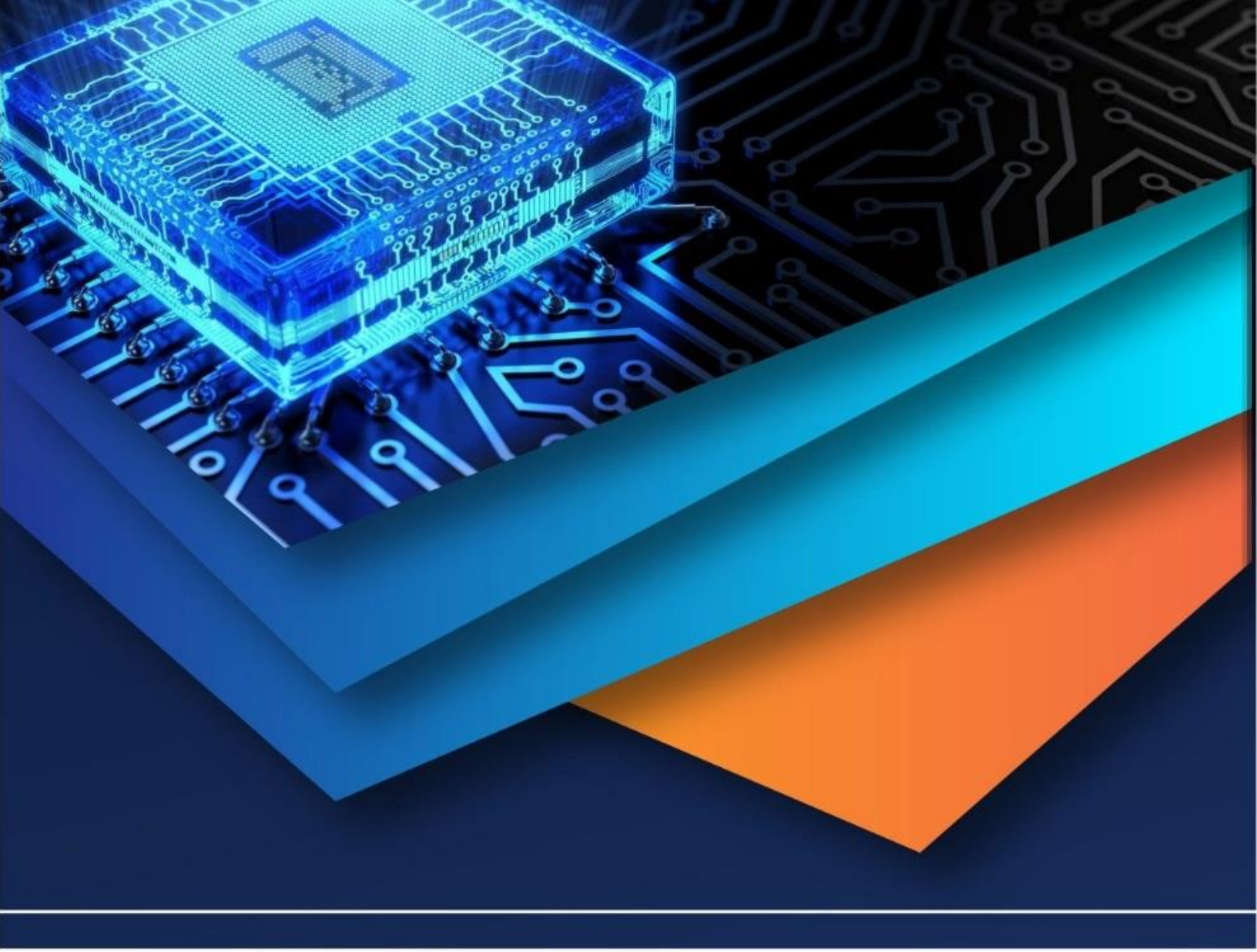
V. CONCLUSION

This paper overviewed the AI Chatbots and the technologies used behind them. It explored the various advancements in technologies such as natural language processing, pattern recognition, semantic web, data mining and transformer-based models that have contributed to the effectiveness of chatbots in diverse industries. This can help researchers develop a better understanding of the current advancements in technologies and usages of chatbots. Based on the chatbot market revenue statistics mentioned earlier in the paper, the chatbot market has shown a phenomenal growth since 2018 and it is rapidly growing every year. It is expected to grow exponentially in the future as many chatbots have been introduced in recent years. With every technological advancement, chatbots are becoming more and more human. The chatbot technology has evolved to such a great extent, that it has become difficult to recognize whether we are interacting with a chatbot or a human. As we look to the future, it is clear that chatbots will play an increasingly important role in shaping the way we communicate and do business.

REFERENCES

[1] Abu Shawar BA. A Corpus Based Approach to Generalise a Chatbot System [PhD thesis]. University of Leeds, Leeds; 2005.
[2] https://en.wikipedia.org/wiki/Chatbot
[3] Systematic Review on Chatbot Techniques and Applications Dong-Min Park, Seong-Soo Jeong, and Yeong-Seok Seo*.
[4] Chatbots: Security, privacy, data protection, and social aspects Martin Hasal1 Jana Nowaková2 Khalifa Ahmed Saghair2 Hussam Abdulla2 Václav Snášel2 Lidia Ogiela3.
[5] https://venturebeat.com/ai/facebook-messenger-hits-100000-bots/
[6] https://www.startus-insights.com/innovators-guide/natural-language-processing-trends/
[7] An Overview of Artificial Intelligence Based Pattern Matching in a Security and Digital Forensic Context Faye Rona Mitchell Chapter 17.
[8] Machine Learning Pattern Matching Azhar Ushmani Senior Software Engineer AdvancedMD Inc USA.
[9] Eduard Puerto, Jose Aguilar & Danilo Chávez (2018): A recursive patterns matching model for the dynamic pattern recognition problem, Applied Artificial Intelligence, DOI: 10.1080/08839514.2018.1481593.
[10] https://www.senseforth.ai/conversational-ai/how-do-chatbots-work/
[11] https://en.wikipedia.org/wiki/Semantic_Web
[12] https://www.dataversity.net/semantic-technology-trends-in-2024/ By Keith D. Foote on December 12, 2023.
[13] G. Antoniou and F. V. Harmelen, A Semantic Web Primer, 2nd ed. Cambridge, MA: MIT Press, 2008.
[14] DATA MINING TECHNIQUES AND APPLICATIONS Bharati M. Ramageri / Indian Journal of Computer Science and Engineering Vol. 1 No. 4 301-305
[15] https://www.mdpi.com/topics/Recent_Advances_Data_Mining
[16] Radford, Alec; Narasimhan, Karthik; Salimans, Tim; Sutskever, Ilya (June 11, 2018). "Improving Language Understanding by Generative Pre-Training".
[17] What is GPT AI? - Generative Pre-Trained Transformers Explained - AWS (amazon.com)
[18] Devlin, Jacob; Chang, Ming-Wei; Lee, Kenton; Toutanova, Kristina (October 11, 2018). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding".
[19] https://en.wikipedia.org/wiki/BERT_(language_model)#cite_note-:0-1 \
[20] https://www.geeksforgeeks.org/explanation-of-bert-model-nlp/

[21] https://www.zdnet.com/article/best-ai-chatbot/
[22] https://claudeai.uk/claudeai-pros-and-cons/#The_Cons_ClaudeAIs_Current_Limitations
[23] "Jasper AI Review: The Good, The Bad, and The Robot" Jussi Hyvarinen https://www.linkedin.com/pulse/jasper-ai-review-jussi-hyvarinen-tifyf
[24] https://en.wikipedia.org/wiki/Gemini_(chatbot)
[25] 65 Chatbot Statistics & Trends 2024 (Must know) By Rohit Shewale / March 6, 2024 65 Chatbot Statistics & Trends 2024 (Must know) (demandsage.com)
[26] Modern Chatbot Systems: A Technical Review Abbas Saliimi Lokman(B) and Mohamed Ariff Ameedeen IBM Centre of Excellence, Faculty of Computer Systems and Software Engineering, Universiti Malaysia Pahang, Pekan, Malaysia {abbas,mohamedariff}@ump.edu.my
[27] An Overview of Chatbot Technology Eleni Adamopoulou(B) and Lefteris Moussiades Department of Computer Science, International Hellenic University, Agios Loukas, 65404 Kavala, Greece {eladamo,lmous}@cs.ihu.gr
[28] Brandtzaeg, P.B., Følstad, A.: Why people use chatbots. In: Kompatsiaris, I., et al. (eds.) Internet Science, pp. 377–392. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-70284-1_30
[29] Review of State-of-the-Art Design Techniques for Chatbots Ritu Agarwal1 · Mani Wadhwa1
[30] Van Merri B, Fellow CS. Learning phrase representations using RNN encoder—decoder for statistical machine translation. arXiv preprint arXiv: 1724–1734. 2014.
[31] Li X, Mou L, Yan R, Zhang M. StalemateBreaker : a proactive content-introducing approach to automatic human-computer conversation. arXiv preprint arXiv:1604.04358. 1:2845–2851.
[32] Sutskever I, Vinyals O, Le QV. Sequence to sequence learning with neural networks. In: Advances in neural information processing systems. p. 3104–12.



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)