

# Unveiling the Potential of ChatGPT: Applications, Challenges, and Future Directions

Hiba Mohsin  
Aligarh Muslim University,  
India

Sehba Masood  
Aligarh Muslim University,  
India

## ABSTRACT

ChatGPT is a large language processing model created by OpenAI. It has brought significant improvements to the field of natural language processing, particularly generating text and conversational dialogues as well as answering questions. In this paper, a systematic literature review on ChatGPT has been done. Intelligent approaches and techniques associated with Large Language Models have been discussed. Further, the salient features of ChatGPT including generating human-like responses and understanding natural language have been examined. Moreover, the issues and challenges of ChatGPT such as bias, misinterpretation, security concerns, etc., have been highlighted. Finally, the various applications of ChatGPT across various sectors like healthcare, banking, business, and content generation have been discussed. Overall, this work provides valuable insights for researchers and industries seeking to enhance the performance and application of ChatGPT.

## Keywords

Artificial Intelligence, Natural Language Processing (NLP), Large Language Model (LLM), ChatGPT.

## 1. INTRODUCTION

Artificial intelligence (AI) encompasses the development of algorithms and computer programs that mimic human intelligence, enabling machines to perform several tasks related to speech-to-text conversion, computer vision, and other real-world applications [1]. AI can be classified into two types: strong AI, which is built to accomplish every intellectual task that a human can, and weak AI, which is built to perform only a few specific activities[2]. Artificial intelligence (AI) applications include robots, autonomous cars, image identification, and natural language processing. AI systems use machine learning algorithms to perform large-scale data analysis and interpretation, and decisions are subsequently made in response to the analysis. Deep learning and other recent developments in AI have made tremendous strides in the discipline, allowing for the development of language models like GPT-3 that can produce human-like language and carry out a variety of tasks.

ChatGPT is a sophisticated language model created by OpenAI that makes use of artificial intelligence to respond to inputs in natural language that are human-like. It can comprehend a wide range of topics and is designed primarily for conversational applications like chatbots and virtual assistants. However, it has various difficulties that need to be resolved. The likelihood that the model will produce biased or unsuitable responses is one of the major difficulties. This indicates that the model might generate responses that reflect and support preexisting biases. Another difficulty is its tendency to generate irrelevant or off-topic responses. To overcome these problems, researchers are looking into a variety of approaches, such as expanding the

diversity of training data and enhancing the model's capacity for language generation and comprehension.

The motivation for this work is to present an extensive review of the current state of research on ChatGPT. This paper included its intelligent approaches and techniques. How ChatGPT is being used in various applications. The unique features of ChatGPT were discussed, as well as the challenges it faces and the efforts made to address them to achieve its full potential. Ultimately, this work aims to provide a valuable tool for researchers, developers, and practitioners seeking to enhance their natural language processing tasks through the utilization of ChatGPT.

The absence of comprehensive surveys examining the issues and challenges of ChatGPT and its applications has driven our research motivation. A systematic literature review process was undertaken, following a step-by-step protocol, to identify, select, and evaluate relevant studies that address specific research questions. The study concentrates on reviewing research conducted in this field between 2010 and 2023. To guide the investigation, four primary research questions were formulated as follows:

- Gain a comprehensive understanding of the intelligent approaches and techniques developed about Language Model (LLM) systems.
- Conduct a comparative analysis of the performance of various existing models and techniques of GPT (Generative Pre-trained Transformers) on different aspects.
- Explore the current issues and challenges with ChatGPT and future research directions to overcome them.
- Discuss the various applications of ChatGPT.

The subsequent sections of the paper are organized as follows. Section 2 presents the methodology and guidelines for conducting a systematic literature review. Section 3 showcases the findings of the review and provides a comprehensive discussion addressing the research questions. Finally, Section 4 concludes the review by summarizing the key findings.

## 2. METHODOLOGY

This study employs a systematic literature review approach, adhering to a pre-defined step-by-step protocol, to identify, select, and assess relevant studies that pertain to specific research inquiries. The scope of our investigation spans from 2010 to 2023, with a focus on reviewing research conducted in the field. Firstly, the research seeks to attain an all-encompassing understanding of the various intelligent methodologies and techniques developed in the realm of Language Models. Language Models have been at the forefront of natural language processing, and grasping the intricacies of the advancements in this area is critical to fostering further progress. Through an exhaustive review of relevant literature,

the study endeavors to synthesize and present a cohesive overview of the LLM landscape.

Secondly, the research endeavors to conduct a comparative analysis of the performance of diverse existing GPT models and techniques. Given the proliferation of GPT-based models and architectures, a thorough evaluation of their respective capabilities and limitations becomes indispensable. By analyzing the outcomes of different GPT variants across multiple facets, this study aims to provide valuable insights for researchers and practitioners, facilitating informed decisions in choosing appropriate models for specific tasks.

Thirdly, the study sets its sights on exploring the prevailing issues and challenges that surround ChatGPT, an application of GPT that has gained substantial popularity. While ChatGPT holds tremendous promise in various interactive settings, it also faces certain hurdles, such as generating biased or inappropriate responses and adhering to ethical guidelines. The research endeavors to shed light on these challenges and proposes potential research directions to tackle these issues effectively, ensuring responsible and safe use of such language models.

Finally, the research aims to explore and discuss the manifold applications of ChatGPT in real-world contexts. As language models continue to evolve, their versatility in diverse applications becomes evident. From customer service chatbots to educational tools and creative writing assistance, ChatGPT has found practical utility in numerous fields. Through a meticulous review of existing literature, this study endeavors to provide a comprehensive overview of the applications and implications of ChatGPT, furthering our understanding of its potential impact.

## 2.1 Research questions

A systematic review was conducted to identify, select, and critically evaluate relevant academic works focusing on GPT. The evaluation was performed using multiple criteria, as outlined in the results section. To enhance the organization of our findings, we formulated four primary research questions as follows:

- RQ1-What are the intelligent approaches and techniques developed related to LLM?

Motivation: This question focuses to provide a comprehensive overview of the intelligent approaches and techniques developed related to LLMs based on GPT, highlighting their importance in improving performance and applicability and providing insights for future research in this rapidly growing field.

- RQ2- What are the salient features of ChatGPT that make it an effective LLM?

Motivation: This question aims to provide a brief knowledge of its unique features and capabilities making it an ideal choice for businesses and individuals who are looking to improve their communication and customer service capabilities.

- RQ3- What are the current issues and challenges with ChatGPT and ongoing efforts to address those issues and challenges?

Motivation: The main objective of this question is to find out the problems and difficulties that need extra attention and work to improve the performance of the ChatGPT system. By understanding these challenges and the ongoing efforts made to solve them, we can make the system better.

- RQ4- What are the applications of ChatGPT?

Motivation: By exploring the various applications of ChatGPT, we can gain a better understanding of its potential impact on various domains and identify new opportunities for research and development.

As shown in Table 1, A list of search strings or phrases, each serving a specific purpose related to ChatGPT, was compiled for searching papers in various databases or web browsers. Each search string was carefully crafted with a particular motive in mind.

**Table 1. Shows the specific words or phrases that were used during a random search in different sources or web browsers to find a particular research paper**

Specific words/Phrases	Expected Search Results
“ChatGPT”	Sources with the string “ChatGPT” string
"Generative Pre-Trained Transformer"	A source with the term "Generative Pre-Trained Transformer"
“ChatGPT AND Large Language Models”	Sources having the ChatGPT and Large Language Models
“ChatGPT And Applications”	Sources having all the terms ChatGPT and Applications
“ChatGPT issues And Challenges”	Sources with the terms ChatGPT issues and Challenges

As shown in Table 2, A list of specific journals and conference proceedings was created, carefully selected to search for research papers specifically related to ChatGPT. Every item in the list has an abbreviation, the journal or conference proceedings name, and the publication source.

**Table 2. Consists of a chosen set of journals and conference proceedings that were taken into account for the methodical search for papers**

Abbreviation	Journal Name	Publication Source
JAIR	1. Journal of Artificial Intelligence Research	AI Access Foundation
IEEE TNNLS	2. IEEE Transactions on Neural Networks and Learning Systems	IEEE
ACM TIS	3. ACM Transactions on Interactive Intelligent Systems	ACM
ACL AM	4. Association for Computational Linguistics (ACL) Annual Meeting	ACL
IEEE TPAMI	5. IEEE Transactions on Pattern Analysis and Machine Intelligence	IEEE
AI	6. Artificial Intelligence	Elsevier
MI	7. Machine Learning	Springer
Conference Proceedings		
IJCA	1. International Joint Conference on Artificial Intelligence	IJCAI
ICLR	2. International Conference on Learning Representations	OpenReview
ACL CoNLL	3. Conference on Natural Language Processing	ACL

## 2.2 Search process

Published research papers and conference proceedings between 2010 and 2023 were used as the primary search strategy to get the data. There are three different types of journals in the primary source peer-reviewed publications, conceptual journals, and literature review journals. A total of seven journals and three conference proceedings were selected for the search, as detailed in Table 2. Due to the limitations of automatic searching, a decision was made to manually search for papers. The initial focus was on the journals, utilizing specific criteria such as journal name, year, volume, and issue to narrow down the search within defined periods. However, the system search did not yield a sufficient number of papers, prompting the adoption of a random search approach. Various search terms were employed across different databases and web browsers, including Google Scholar, Scopus, Web of Science, ACM Digital Library, IEEE Explore, and other relevant websites. To determine the relevance of the papers, a comprehensive evaluation was performed on the title, abstract, and complete content of each paper.

## 2.3 Study evaluation criteria

Protocols have been established to ensure the relevance and validity of the selected primary data. These protocols serve as guidelines for assessing the accuracy and appropriateness of the chosen data.

### 2.3.1 Inclusion

- Conference proceedings, articles, and research papers about ChatGPT or its various variants.
- Articles that discuss the creation, evaluation, and application of ChatGPT for different natural language processing (NLP) problems.
- Research analyzing the advantages and disadvantages of ChatGPT by comparing it with alternative language models or baselines.
- Papers that address potential applications of ChatGPT in different domains
- Studies that examine the bias and privacy issues in connection with the use of ChatGPT.

### 2.3.2 Exclusion

- Research that is irrelevant to ChatGPT or its variants.
- Studies that just present ChatGPT benchmark data without offering any qualitative analysis or ideas.
- Articles that do not provide any innovative ideas or contributions to the existing literature on ChatGPT.
- Papers that are not written in the English language.
- Research that has not been published in respected publications or professional journals.

## 2.4 Initial selection process

The initial selection process involves analyzing the relevance of the selected papers. Initially, the relevance is assessed by thoroughly examining the title and abstract. If necessary, a full paper review is conducted. The total number of papers selected after this selection procedure is presented in Table 3.

**Table 3. Procedure for selecting studies**

Step	Selection Guidelines	Number of Research Papers
1.	Total number of papers found on selection guidelines	152
2.	Eliminated based on abstracts and conclusions	97
3.	Selected for primary studies	58

## 2.5 Final selection process

After considering the inclusion and exclusion criteria, a comprehensive examination of the initially selected papers was conducted. Two researchers reviewed the criteria and reached a consensus through detailed discussions. By thoroughly reading the complete content of the papers, any inquiries regarding them were addressed. Ultimately, a total of 58 papers were chosen for the literature review.

### **3. RESULT**

Within this section, the research questions that were selected are evaluated using the data gathered during the systematic literature review (SLR). In the subsequent subsections, each particular question is addressed and answered.

#### **3.1 Intelligent approaches and techniques developed related to LLMs**

GPT and similar Large Language Models have revolutionized natural language processing. Various intelligent approaches and techniques have been developed to improve the performance and application of LLMs based on GPT. This paper presents a comprehensive overview of the intelligent approaches and techniques developed for LLMs based on GPT. The various techniques utilized to train and fine-tune LLMs, including transfer learning and pre-training, are explored. Furthermore, the applications of LLMs in diverse domains are discussed.

In this section, intelligent approaches and techniques developed about Language Models (LLMs) based on GPT were explored, providing an answer to RQ1-What are the intelligent approaches and techniques developed related to LLMs based on GPT?

In 2018, OpenAI introduced the GPT (Generative Pre-trained Transformer)[3] model, which was the first publicly available LLM. This model could produce text that seemed to be written by a human, answer questions, and aid in tasks like summarization and translation using natural language completion. OpenAI subsequently released GPT-2 and GPT-3, which improved upon the capabilities of the initial model. The release of GPT is considered a major accomplishment in the domain of natural language processing, as it has provided researchers and industries with many new ways to use and apply the technology.

In 2018, Google Research introduced BERT, which stands for Bidirectional Encoder Representations from Transformers [4]. Like GPT, BERT is based on transformers and pre-trained on extensive text. BERT undergoes training with two unsupervised tasks, masked language modeling, and next-sentence prediction, enhancing its contextual comprehension. Masked language modeling is a task where the model is given a sentence with missing words, and it must predict what those missing words are based on the context of the sentence. Next sentence prediction involves the model predicting the most likely sentence to follow a given sentence. By training on these tasks, BERT grasps contextual word knowledge across diverse subjects.

Google AI rolled out XLNet in 2019, which uses permutation language modeling to understand the relationships between words in a sentence [5]. This approach makes XLNet suitable for tasks such as natural language inference and question answering. In 2020, Google Research introduced T5[6], a transformer-based model trained on extensive text data, capable of performing multiple NLP tasks. Facebook AI also entered the field of large language models with the development of RoBERTa (Robustly Optimized BERT Pre-training)[5]. Roberta, an enhanced version of BERT, employs dynamic masking in pre-training and was trained on a larger dataset. Upon release, Roberta surpassed BERT, GPT-2, and XLNet in performance.

BLOOM, a new large language model by BigScience, was developed collaboratively and trained on a diverse dataset of 46 natural languages and 13 programming languages, totaling

1.6TB[8]. It aims to provide an openly accessible and transparently trained multilingual model for academic and nonprofit use.

OpenAI's GPT (Generative Pre-trained Transformer) models are powerful neural language models based on the Transformer architecture. They generate natural language text that resembles the input text in style and content. GPT-3 is the current largest and most widely used language model. Its massive pre-training dataset includes books, articles, and websites, among others, with 175 billion parameters. It is also a transformer-based model to generate text of exceptionally high coherency. The text produced by GPT-3 closely resembles human-written content to such an extent that distinguishing between them becomes challenging [6]. GPT-3 has a remarkable capacity for zero-shot learning, enabling it to perform tasks for which it has not specifically been trained. This characteristic unlocks tremendous opportunities for a wide range of applications, including automation, summarization, creative writing, chatbots, and dialogue systems.

GPT-3.5 is a variant of the GPT-3 model that was developed by OpenAI to address some of the issues with the original GPT-3 model, such as the potential for generating biased or offensive text. It incorporates additional training data and techniques to improve the fairness and inclusivity of the language generated by the model.

The popular conversational AI model, ChatGPT, is built on the GPT architecture. It is specifically designed for generating coherent and engaging text-based conversations with humans. ChatGPT has shown promising results in various conversational tasks such as chatbot development, customer service, and virtual assistants. Training on extensive conversational data enhances its capability to produce suitable and captivating replies to human input.

Table 4 represents the comparison of LLMs based on their parameters, pretraining data, and advantages.

#### **3.2 Salient features of ChatGPT that make it an effective LLM**

ChatGPT is a conversational AI model developed by OpenAI that has gained widespread attention for its remarkable performance in generating natural language text. In this section, the salient features of ChatGPT that make it an effective Large Language Model (LLM) are explored to provide an answer to RQ2- What are the salient features of ChatGPT that make it an effective LLM?

ChatGPT incorporates innovative techniques such as attention mechanisms and transformer networks, which allow the model to process sequential data and generate text that is contextually relevant and coherent. These techniques enable ChatGPT to understand the nuances of human language, such as idiomatic expressions, sarcasm, and humor, which makes it particularly effective in conversational settings.

The model's ability to generate fluent and coherent text, its use of innovative techniques, and its ethical considerations make it a promising tool for various natural language processing applications, including chatbot development, customer service, and virtual assistants.

The large language model called ChatGPT, heavily relies on NLP. ChatGPT can comprehend text-based communication and provide responses that resemble those of a human. By utilizing NLP approaches, ChatGPT can comprehend and analyze conversational text, determine the purpose and context of user messages, and produce appropriate responses in

response to that analysis. The foundation of ChatGPT is deep learning methods, notably transformers, which have shown outstanding results in NLP applications like language modeling, question answering, and text production. By using these methods, ChatGPT can learn from a significant amount of text input and provide meaningful, cogent, and contextually appropriate responses that resemble those of a human. Furthermore, ChatGPT makes use of NLP methods like sentiment analysis and named entity recognition to comprehend user intent and produce more precise responses[7]. Using these methods, ChatGPT can recognize significant entities and themes in the user's message and produce responses that are specific to those things and topics. In general, NLP's function in ChatGPT is to provide it the ability to comprehend

and produce natural language responses to text-based discussions in a way that is human-like and acceptable for the situation.

ChatGPT is a chatbot application that creates human-like responses to text-based discussions using a large language model known as GPT (Generative Pre-trained Transformer). One of the top artificial intelligence research organizations, OpenAI, has created ChatGPT. To produce language that is coherent and appropriate for the context, the GPT deep learning algorithm is trained on a sizable amount of text data[8]. The algorithm is first fine-tuned on specialized tasks like text categorization, summarization, and conversation production before being pre-trained on sizable corpora of text data.

**Table 4 Comparison of Large Language Models**

<b>Model</b>	<b>Parameters</b>	<b>Released</b>	<b>Pertaining Data</b>	<b>Advantages</b>
<b>GPT</b>	110 million	OpenAI in 2018	Web text	-Produces text that looks human-written. -Answer questions and help in tasks like summarization and translation. -First scaled-up transformer model
<b>BERT</b>	340 million	Google Research in 2018	Books, Wikipedia	- The bidirectional transformer model introduced the concept of masked language modeling. -Pre-trained on an enormous data set and improved the knowledge of phrases and words in the content.
<b>GPT-2</b>	1.5 million	OpenAI in 2019	Web text	-Improved Transformer Model
<b>RoBERTa</b>	355 million	Facebook AI in 2019	Internet text, Books, CC-News, Wikipedia	-Improved BERT Model -performed superior because it was trained on a larger dataset than BERT
<b>XLNet</b>	340 million	Google AI in 2019	Books, Wikipedia	-Improved Transformer Model -Uses sequential language modeling to understand how words are related in a sentence. -Ideal for activities like question-answering and natural language processing.
<b>T5</b>	11 billion	Google Research in 2020	Web text, Books, CC-News, Wikipedia, TBC	-Text-to-Text Transformer Model -Perform multiple NLP tasks
<b>BLOOM</b>	176billion	Big Science Community in 2022	A data set consisting of 46 natural languages and 13 programming languages was collected	-GPT3- based multi-lingual corpus
<b>GPT-3</b>	175 billion	OpenAI in 2020	Web text, Books, Wikipedia	- An exceptionally large model was developed, demonstrating the ability for few-shot learning.

ChatGPT produces responses to text-based conversations using the GPT algorithm. When a user delivers a message to ChatGPT, the algorithm examines the text, determines the message's intent and context, and then produces a response that is pertinent and cogent. Frequently, it is impossible to tell the difference between ChatGPT's output and human-generated

responses[9]. Many applications, including chat-based games, personal assistants, and customer support, can be made use of with ChatGPT. It is created to be user-friendly and can be implemented into a variety of platforms and applications to offer a seamless chatbot experience.

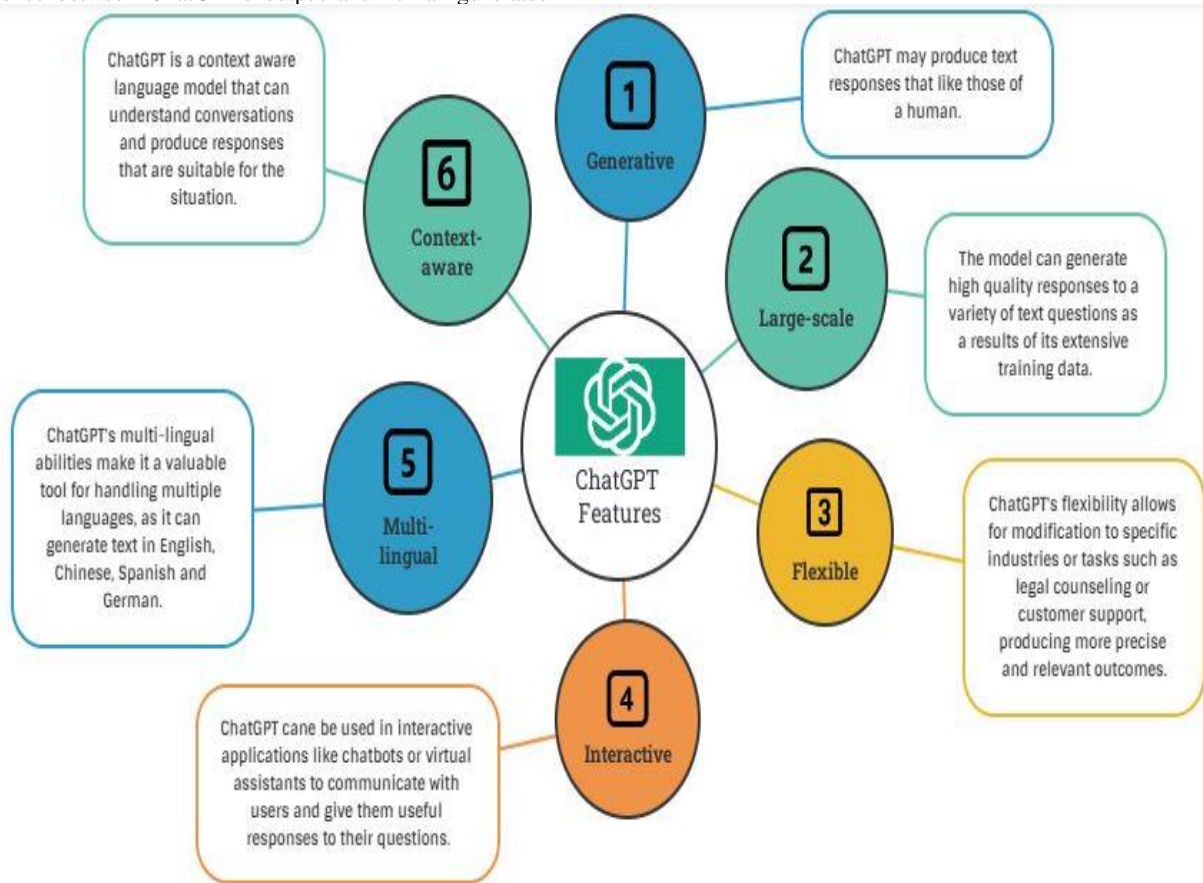


Figure 1 Salient Features of ChatGPT

ChatGPT, developed by OpenAI, is a state-of-the-art language model that has several key features that make it unique and powerful. Here are some of the notable features of ChatGPT:

- **Generative:** Being a generative model, ChatGPT can produce text responses that closely resemble human-like writing when given a specific text prompt [10].
- **Large-scale:** The enormous amount of data, which included billions of words, was used to train ChatGPT. The model can produce high-quality responses to a variety of text prompts because of the extensive training data[11].
- **Context-aware:** ChatGPT is context-aware, which enables it to comprehend conversations and produce responses that consider that context. This enables ChatGPT to produce more logical and pertinent responses to a given text query[12].
- **Flexible:** ChatGPT can be tailored for particular jobs or industries, including giving legal counsel or customer support. Its adaptability enables the model to be tailored for certain use cases and to produce more pertinent and accurate results[13].
- **Multi-lingual:** The languages that ChatGPT can produce text in include English, Chinese, Spanish, French, and German. ChatGPT is a handy tool for

programs that need to handle multiple languages due to its multilingual capabilities[14].

- **Interactive:** ChatGPT enables user communication through interactive conversations. It provides them with useful answers to their questions in interactive apps like chatbots or virtual assistants[15].

These features make ChatGPT an effective language model that can be applied to a variety of tasks, from customer service to creative writing.

There are several benefits of ChatGPT including:

- **24/7 Availability:** ChatGPT is always available to provide assistance or answer questions, regardless of the time of day or night[16].
- **Fast Response Time:** ChatGPT can provide quick responses, without the need for human intervention, which can save time and increase productivity[17].
- **Large Knowledge Base:** ChatGPT has undergone extensive training on a large volume of data, enabling it to offer precise and pertinent information across a wide range of topics. [18].
- **Cost-effective:** For businesses dealing with a large influx of customer inquiries, employing ChatGPT can offer a more cost-efficient alternative to hiring and training human support staff [19].

- **Personalized Responses:** ChatGPT can analyze the context and history of a conversation to provide more personalized responses[20].
- **Scalability:** ChatGPT can handle multiple conversations at once, making it ideal for businesses with high volumes of customer inquiries[21].

Overall, ChatGPT offers quick, reliable, and individualized assistance and service, which can reduce waiting times, boost output, and enhance client satisfaction.

### 3.3 Current issues and challenges of ChatGPT

ChatGPT has made significant strides in natural language processing tasks such as text and conversation generation and question-answering. However, like any advanced technology, ChatGPT is not without its challenges and limitations. In this section, the current issues and challenges faced by ChatGPT are explored, encompassing areas such as bias in language generation, scalability, security, and ethical concerns. The impact of these challenges on the development and utilization of ChatGPT is discussed, along with suggested potential solutions to address these limitations. The importance of ongoing research and development to enhance its performance and application is emphasized to address RQ3- What are the current issues and challenges with ChatGPT and ongoing efforts to address those issues and challenges?

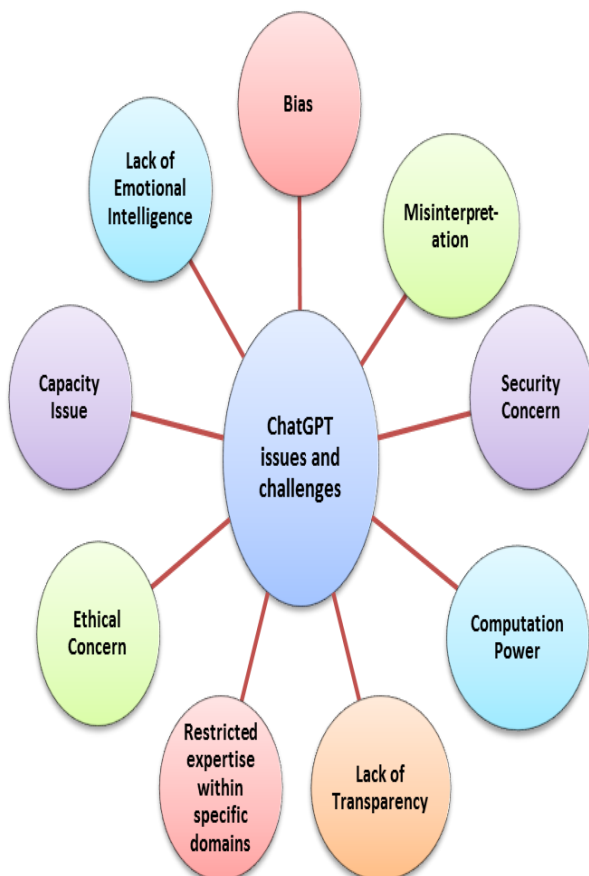


Figure 2 ChatGPT issues and challenges

Next, the current issues and challenges of ChatGPT are presented and future directions to overcome those by one in detail:

#### 3.3.1 Bias

ChatGPT can perpetuate biases present in the training data, which may lead to inaccurate or discriminatory responses [22]. Addressing this requires conscious efforts to ensure diversity and inclusivity in the data used to train AI language models[23]. Bias can manifest in many ways in ChatGPT. For example, the language model may exhibit gender bias, providing different responses to inputs based on the gender of the person interacting with it. It may also exhibit racial or ethnic bias, providing different responses based on the race or ethnicity of the person interacting with it. These biases can be subtle and difficult to detect but can have a significant impact on the users who interact with ChatGPT. Bias can have a large impact on ChatGPT. For instance, if the language model displays gender bias, it can give various answers to questions about jobs depending on the gender of the person asking the question. As a result, opportunities for people depending on their gender may be restricted and stereotypes may be reinforced. Similarly to this, if the language model demonstrates racial or ethnic prejudice, it may offer various answers to questions about healthcare or financial services, resulting in inequities in access to crucial resources.

A proactive strategy is required to solve the bias issue in ChatGPT. Including diversity and inclusion into the training data, actively identifying and resolving algorithmic biases, and continuously monitoring the language model to identify and fix any biases that may develop over time are a few examples of how to achieve this[9]. It might also entail talking to a variety of stakeholders to get their opinions, then incorporating those opinions into the continuing development and enhancement of ChatGPT.

#### 3.3.1.1 Elements Influencing Bias in Large Language Models

When it comes to advanced language models like GPT-4[24, 25], GPT-3, GPT-2, and BERT, [3, 26, 27] as well as other cutting-edge alternatives [28], bias can be described as consistent inaccuracies, faulty attributions, or distortions of facts that show favoritism towards specific groups or ideas. These biases can promote prejudices or lead to false inferences based on observed patterns. Several factors contribute to these biases, including biased training data, algorithmic design, labeling, and annotation processes, product design decisions, and regulatory considerations.

One factor that can lead to bias in language models is the training data itself. These biases may be brought from either the source material used for training or the method used to choose the data. The model gains knowledge from this data, so its behavior could exhibit biases [29-31]. Additionally, Algorithms used for data processing and learning may also incorporate biases. If certain attributes or data points are given more importance by the algorithm, it can inadvertently create or amplify existing biases in the data [32, 33]. Human annotations that add labels or annotations to the training data may also introduce biases since their arbitrary judgments may affect how well the model understands the data [34, 35].

The design and development of large language models can lead to biases in various ways. One of the factors is the prioritization of certain use cases or the user interface design, which can unintentionally reinforce existing biases and neglect diverse perspectives[36]. Another factor is the influence of policies that

guide the behavior of language models. Developers can establish guidelines that aim to restrict or promote particular behaviors of language models. To address potential issues, OpenAI and Microsoft have implemented protective measures to minimize harmful behaviors and safeguard against the malicious misuse of their language models, like ChatGPT and Bing-AI. These guidelines serve as barriers to ensure that the models' actions are not toxic and cannot be exploited for malicious purposes [37, 38].

### 3.3.1.2 Types of Biases in Large Language Models

Large language models are often trained using massive amounts of text data from the Internet. As a result, they unavoidably absorb the biases that exist within that data. These biases might manifest in the following ways:

- **Demographic biases** happen when the training data either give too much or too little representation of certain demographic groups. As a result, the model may exhibit biased behavior towards those genders, racial or ethnic groupings, or other social groups [39].
- The presence of cultural biases in the data used for training large language models can lead to the perpetuation of cultural stereotypes or prejudices. As a result, the model may generate outputs that reinforce or worsen existing cultural biases [40].
- Linguistic biases can occur because the majority of online content is written in English or other widely spoken languages. This gives large language models a higher level of proficiency in these languages, resulting in biased performance. As a consequence, low-resource languages or minority dialects may not receive adequate support from these models [41, 42].
- Temporal biases can arise in language models due to the limited temporal scope of the training data or temporal cutoffs in the data. The model may exhibit biased reporting of events, trends, and opinions. Additionally, its comprehension of historical contexts or accuracy might be limited due to insufficient representative data [3].
- Confirmation biases can be present in the training data of language models because people tend to look for information that confirms their pre-existing beliefs. So, LLMs can unintentionally perpetuate biases by generating outputs that affirm certain concepts. [4].
- Lastly, large language models are vulnerable to ideological and political biases because these biases exist in the data used to train them. As a result, the model's biased outcomes can amplify existing biases towards certain ideologies or viewpoints [43, 44].

### 3.3.1.3 What makes creative language models susceptible to bias?

#### i. Data-driven biases

ChatGPT and similar applications that utilize large language models are primarily trained through unsupervised learning. This means that the models learn to identify patterns and structures from a massive amount of unlabeled data [45]. Typically, these language models are trained using extensive text collections from various online sources including web pages, papers, novels, and other written sources.

ChatGPT has undergone training using a wide array of internet text datasets, encompassing different subjects, styles, and languages. While the precise details of the dataset utilized for GPT-4 are not publicly disclosed, it is expected to be similar to the dataset employed for its predecessor, GPT-3. The main dataset employed for GPT-3 is WebText, which consists of an extensive compilation of web pages obtained by systematically scanning the internet. This dataset comprises sources such as news websites, blogs, forums, e-commerce sites, social media platforms, educational websites, government websites, and online encyclopedias. Besides WebText, GPT-3 was trained on other datasets such as Common Crawl, Wikipedia, and BookCorpus, among others. These datasets cover diverse topics and languages, enabling the model to learn from a broad range of texts. The size and composition of these datasets may differ among models and versions.

- **Web pages:** Text is collected from various types of web pages, such as news sites, blog posts, forums, and educational platforms like Encyclopedia. This allows the model to learn from a diverse range of sources and gather knowledge on a wide range of subjects.
- **Books:** The training material consists of a diverse collection of text sources, including online books of various genres such as fiction as well as non-fiction. This enables the model to pick up a lot of knowledge from many different subjects as well as varied writing styles and narrative patterns.
- **Social Media Platforms:** To provide the model with exposure to informal language, slang, and current topics of discussion, text from social media platforms such as Twitter, Facebook, and Reddit is included in the dataset.
- **Conversational Data:** To make the model better at having conversations, the training data includes text from chatting, commentary, and other sources where people talk to each other online. This helps the model learn how to engage in natural and interactive discussions.

According to ChatGPT developers, before the WebText data is fed into the model, it is preprocessed and filtered to remove spam, offensive language, and other undesired text [46]. Unfortunately, due to the vast amount of data and the current limitations of filtering methods, there is a possibility that the training dataset may contain undesirable or biased information. To train GPT-3, a modified version of the Common Crawl dataset was utilized. Common Crawl is a publicly accessible dataset obtained by crawling the web, containing a wide range of online page data, including metadata and textual information from billions of web pages in various languages. Moreover, the WebText dataset was also employed in the training process [27].

The Pile [47] is a large and diverse compilation of 22 smaller datasets, bringing together various sources such as academic journals, novels, and online content. It is frequently utilized as a dataset for training language models. Particularly, it is beneficial for training large-scale language models that specialize in scientific research and understanding.

#### ii. Biases originating from the models

ChatGPT and similar generative language models undergo training with extensive text data to grasp the interplay between words, grammar, context, and meaning. This helps them generate responses that make sense and fit the conversation.



However, these models can unintentionally develop biases, which can create unfairness and inequality in the AI system.

- **Generalization:** A major concern lies in the generalization capability of such models, as they can produce contextually suitable responses and predictions. However, this ability also gives rise to worries regarding potential biases. Even with meticulous data filtering, these models may unintentionally acquire and reinforce biases from their training data [30, 48].
- **Propagation:** Propagation is another issue that can lead to biased models. When these models learn from the trends and frameworks found in their training data, there is a risk that they might unintentionally acquire and perpetuate biases they encounter. This could include accepting preconceptions or demonstrating preferences toward particular groups or ideas [49].
- **Emergence:** Large language models can also have unexpected biases that emerge on their own, even though these abilities were not explicitly programmed into the model's design or training data. These unexpected biases can occur because of how the model is set up and the biased information it learned during training [50].
- **Non-Linearity:** Non-linearity is another problem that arises because of the complicated connections between the various settings of the model and the complex patterns it learns during training. Even small biases can have significant negative impacts, while large biases may not always lead to noticeable consequences. This makes it difficult to foresee or manage these biases.
- **Alignment:** To tackle these problems, researchers are actively investigating ways to reduce biases during the training process. They are exploring strategies such as incorporating bias mitigation techniques, fine-tuning models using carefully selected datasets, and conducting thorough analyses to identify and address emergent biases after the training phase. Reducing biases and guaranteeing alignment with human values are the goals of the RLHF (Reinforcement Learning with Human Feedback) technique, which is used to enhance big language models like ChatGPT. This approach gathers a dataset of human contrasts, presentations, and preferences, and a reward model and a reward model are then created using the dataset. To improve the performance of the model, this reward model serves as a guide for tuning.

### *3.3.2 Misinterpretation*

ChatGPT can provide responses based on the information it has been trained on, which can sometimes include misinformation or fake news. This can be problematic as it can lead to users receiving inaccurate information and beliefs, which can have negative consequences [18]. This may occur for many reasons, including ambiguity in the user's language, a lack of context, or cultural differences. For instance, if a user asks a question with irony or sarcasm, ChatGPT can view the query as serious and send out an inappropriate response. Similarly, if a user's wording is imprecise or unclear, ChatGPT may produce a response that is unrelated to the user's real query.

### *3.3.3 Security Concern*

As an AI language model, ChatGPT presents several security risks that need to be carefully considered and addressed. These risks can arise from the model's ability to generate text based on its training data, which can be manipulated to produce malicious or harmful outputs.

One of the primary security risks associated with ChatGPT is the potential for the model to be used for phishing attacks or social engineering. Malicious actors could use the model to generate convincing messages or emails that appear to come from legitimate sources, such as financial institutions or government agencies, to trick users into divulging sensitive information or carrying out harmful actions[51]. This could lead to significant financial loss, reputational damage, or other harm.

Another security risk related to ChatGPT is the potential for the model to be used for impersonation or identity theft purposes[52]. The model's ability to generate text that is indistinguishable from human-written text could be used to impersonate individuals online, creating fake social media profiles or other online personas for malicious purposes. This can have serious implications for individuals whose identities may be stolen or manipulated, leading to reputational damage, financial loss, and other harm. Businesses, governments, and people whose identities might be stolen or controlled could be seriously impacted by this.

Moreover, ChatGPT's ability to generate text based on its training data can also be manipulated to produce false or misleading information. Malicious actors could use the model to generate false news stories or social media posts that are designed to spread misinformation or manipulate public opinion [53]. This can have serious implications for democracy and public trust in institutions.

The training data used to train the model can also pose security risks. If the data used to train the model is compromised, it can lead to the model being trained on malicious inputs, resulting in the generation of malicious text outputs[54]. This can be particularly problematic in situations where the model is used to generate text in high-risk environments, such as in the medical or legal fields.

Another security risk associated with ChatGPT is the potential for the model to be attacked by malicious actors seeking to manipulate its outputs [20]. This can be done through various techniques, such as adversarial attacks, where inputs are specifically crafted to mislead the model, or through poisoning attacks, where the training data used to train the model is intentionally manipulated to generate biased or malicious outputs.

### *3.3.4 Computation Power*

Large language models require significant amounts of computation power and resources, making them challenging to train and deploy [55]. Moreover, the energy consumption of these models is substantial, leading to environmental concerns.

### *3.3.5 Lack of Transparency*

Large language models are often considered black boxes as they are difficult to interpret and understand [56]. This lack of transparency can lead to challenges in auditing and validating the results generated by these models.

### 3.3.6 Restricted expertise within specific domains

Large language models are trained on general-purpose datasets, which can result in limited domain-specific knowledge. This challenge can be addressed by fine-tuning these models on specific domains to improve their performance[57].

### 3.3.7 Ethical Concerns

Large language models can raise ethical concerns, including generating harmful content. Ethical guidelines are crucial for developing and deploying these models. The guidelines aim to address concerns related to offensive or harmful output. Establishing ethical standards is vital to ensure responsible model usage[35].

### 3.3.8 Capacity Issue

One of the primary capacity issues with ChatGPT is the amount of computational resources required to generate responses. The model relies on powerful computing systems to process large amounts of data and generate high-quality responses. As a result, there may be times when the model is unable to respond to user queries in real time due to high demand. Another capacity issue with ChatGPT is the size of the model itself. The most recent iteration of ChatGPT, known as GPT-3, boasts an unprecedented scale with over 175 billion parameters [27]. While this allows the model to generate highly accurate and nuanced responses, it also means that the model requires significant storage and processing power.

Lastly, there may be capacity issues related to the scope of knowledge that ChatGPT has been trained on. While it has been trained on a diverse range of topics, there may be areas where the model's knowledge is limited or outdated. In such cases, the model may provide inaccurate or incomplete responses[58].

### 3.3.9 Lack of Emotional Intelligence

ChatGPT does not have emotional intelligence in the same way that humans do[59]. Emotional intelligence involves the ability to perceive, understand, and respond to emotions in oneself and others, and it is a key aspect of human communication and interaction. While ChatGPT can provide accurate and informative responses to user queries, it is not capable of understanding the emotional context of a conversation or empathizing with the user's feelings. This means that ChatGPT may not always provide the emotional support or understanding that a human might provide in a similar situation [60].

Overall, the issues and challenges associated with large language models are significant but can be addressed through appropriate measures.

## 3.4 Applications across various domains

In this section, the various applications of ChatGPT are highlighted and provided an answer to RQ4-What are the applications of ChatGPT

ChatGPT is an incredibly useful tool that can do much more than just scientific research. It has the power to revolutionize different industries, make communication better, and encourage new ideas. Let's take a look at some of the exciting ways ChatGPT can be used in various fields to bring about positive changes.

### 3.4.1 Health services and Medication

In the health services and medication sectors, ChatGPT can be used for various purposes:

- **Assisting doctors:** It can help physicians in identifying diseases by examining patient

information, such as medical history, symptoms, and test results. Additionally, it can create personalized treatment plans based on each patient's specific needs and preferences.

- **Summarizing medical research:** ChatGPT can summarize and synthesize complex medical studies and research papers. This helps healthcare professionals stay updated with the latest evidence-based practices and treatments.
- **Providing patient information:** It can offer medical information and guidance to patients in a way that is easy to understand. This can help patients make informed decisions about their health and treatment options.
- **Enhancing collaboration:** ChatGPT facilitates communication and information sharing among healthcare professionals, promoting collaboration and teamwork. It streamlines the exchange of vital data and insights, leading to improved patient care[20].

There are several possible uses of ChatGPT in healthcare and medicine, which are as follows:

- a. **Helping with patient triage:** ChatGPT can create chatbots that assist healthcare workers in determining how urgently a patient needs care and suggesting appropriate actions[61].
- b. **Providing diagnostic and treatment help:** ChatGPT offers medical professionals helpful diagnostic and treatment recommendations based on patient data and symptoms. [62].
- c. **Supporting medical science education:** ChatGPT aids medical science education by sharing medical information and treatment options [63].
- d. **Assisting in clinical data analysis:** ChatGPT's analysis of extensive clinical data uncovers patterns, aiding treatment innovation [64].

These applications demonstrate how ChatGPT can improve healthcare outcomes, aid in medical decision-making, and advance medical research and education.

### 3.4.2 Finance and Business

ChatGPT can be used in different business and finance domains, showcasing its effectiveness in improving efficiency, productivity, and overall performance in the following ways:

- **Automation of financial reports and market analyses:** ChatGPT can be utilized to automatically produce financial reports and summaries of market analyses, saving time and effort for businesses.
- **Sentiment analysis for product development and marketing:** By analyzing the reviews and feedback provided by customers, ChatGPT can gather valuable information about their experiences and opinions. This data can be utilized to improve products and services, as well as to enhance customer satisfaction and engagement.
- **Individualized investment suggestions:** ChatGPT provides individualized investment guidance by taking into account an investor's risk tolerance and financial goals. This helps investors receive valuable suggestions and recommendations that cater specifically to their needs.

- **Support for written content creation:** ChatGPT can assist with the creation of project proposals, promotional materials, and other written documents, making the process more efficient and effective.
- **Enhanced customer service:** ChatGPT can quickly and accurately respond to consumer inquiries, improving customer service operations by providing appropriate and contextually relevant answers.

These applications demonstrate how ChatGPT can bring automation, data-driven insights, and improved efficiency to the business and finance industry.

ChatGPT has several corporate and financial applications, including customer service chatbots, market analysis and forecasting, financial management support, fraud detection, and risk management. It can provide product recommendations, analyze financial data, identify trends, and help make informed investment decisions while detecting fraud and supporting risk mitigation [67-70]. These applications highlight how ChatGPT can improve customer interactions, provide insights for decision-making, and enhance risk management in the business and finance sector.

### 3.4.3 Content Creation and Creative Writing

In the field of creative writing and content creation, ChatGPT can be utilized for various purposes:

- **Generating story ideas and characters:** ChatGPT can help come up with unique and engaging ideas for stories, provide summaries of plots, and describe characters in detail.
- **Assisting with writing challenges:** ChatGPT can offer creative suggestions and directions to help writers when they are stuck or experiencing writer's block.
- **Automated content creation:** ChatGPT can automatically generate content for blogs, articles, and social media posts based on specific input criteria and preferred writing style, making it easier to generate content for various platforms.
- **Editing and proofreading:** ChatGPT can be used to review and improve written material, ensuring proper grammar, clarity, and coherence.
- **Summarizing written resources:** ChatGPT can create captivating and informative summaries of written materials, providing condensed versions of books, articles, and other resources.

To create content and write creatively, ChatGPT can be helpful in the following ways:

- Create content:** ChatGPT may provide content for a variety of uses, including blogging entries, developments on social networks, and promotional materials. It can provide enlightening and engaging commentary by analyzing the topic, tone, and style [65].
- Writing ideas:** If authors are struggling to generate new ideas, ChatGPT can help by providing unique and imaginative writing prompts. By examining information on genres, themes, and narrative structures, ChatGPT can offer writers fresh perspectives and innovative concepts to inspire their writing. This can spark novel thoughts and approaches to storytelling [66].
- Writing novels:** For writers working on their novels, ChatGPT serves as a valuable asset. It offers valuable

aid by suggesting ideas and providing guidance on plot advancement, character evolution, and the overall framework of the story. By examining data on prevalent genres and narrative structures, ChatGPT provides customized support to authors, enabling them to fashion captivating and enthralling tales.

- Screenwriting assistance:** ChatGPT is a helpful resource for people attempting to break into the screenwriting industry. It provides practical suggestions and creative insights for crafting compelling stories, developing memorable characters, and structuring narratives effectively. By drawing upon its analysis of prevalent genres and story frameworks, ChatGPT offers personalized guidance to writers, empowering them to create captivating and appealing scripts.
- Songwriting support:** ChatGPT serves as a valuable tool for songwriters seeking assistance. It provides practical advice and creative ideas for crafting melodies and lyrics. ChatGPT provides individualized advice to help songwriters develop songs that connect with listeners by evaluating a large amount of data on popular musical genres and topics [20].

These applications highlight how ChatGPT can support writers, enhance content creation, and improve the overall writing process.

## 4. CONCLUSION

ChatGPT is a large language model created by OpenAI that makes use of artificial intelligence to respond to the inputs in natural language that are human-like. It can understand a wide range of topics. Particularly designed for conversational applications like chatbots and virtual assistants. To ensure its equitable and inclusive use major issues and challenges must be resolved. The biases in these models, which might result in inappropriate and discriminatory responses are one of the major issues. By addressing the challenges encountered by ChatGPT, advancements can be made in the field of natural language processing, promoting ethical and responsible utilization of this technology. A systematic literature review (SLR) was conducted on ChatGPT. Kitchenham's guidelines are followed in this SLR. This paper presents a state-of-art study on ChatGPT, which discusses intelligent approaches and techniques related to LLMs with a comparison of LLMs. The salient features of ChatGPT including multi-lingual, context-aware and generative, etc. have also been discussed which makes it unique. The applications of ChatGPT across various domains are also discussed.

Further research and development can focus on improving the continued efforts in addressing ethical considerations, such as minimizing bias and ensuring transparency, which will be essential for the responsible deployment of ChatGPT across various domains. The future of ChatGPT holds the enormous potential to transform communication and improve human-machine interactions with continued developments and interdisciplinary partnerships.

## 5. REFERENCES

- [1] Ergen, M., *What is artificial intelligence? Technical considerations and future perception*. Anatolian J. Cardiol, 2019. 22(2): p. 5-7.
- [2] Burns, E., N. Laskowski, and L. Tucci, *What is artificial intelligence*. Search Enterprise AI, 2021.

- [3] Radford, A., et al., *Improving language understanding by generative pre-training*. 2018.
- [4] Devlin, J., et al., *Bert: Pre-training of deep bidirectional transformers for language understanding*. arXiv preprint arXiv:1810.04805, 2018.
- [5] Liu, Y., et al., *Roberta: A robustly optimized bert pretraining approach*. arXiv preprint arXiv:1907.11692, 2019.
- [6] Clark, E., et al., *All that's human is not gold: Evaluating human evaluation of generated text*. arXiv preprint arXiv:2107.00061, 2021.
- [7] Baidoo-Anu, D. and L. Owusu Ansah, *Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning*. Available at SSRN 4337484, 2023.
- [8] Susnjak, T., *ChatGPT: The End of Online Exam Integrity?* arXiv preprint arXiv:2212.09292, 2022.
- [9] Kasneci, E., et al., *ChatGPT for good? On opportunities and challenges of large language models for education*. Learning and Individual Differences, 2023. **103**: p. 102274.
- [10] Khalil, M. and E. Er, *Will ChatGPT get you caught? Rethinking of plagiarism detection*. arXiv preprint arXiv:2302.04335, 2023.
- [11] Qin, C., et al., *Is ChatGPT a general-purpose natural language processing task solver?* arXiv preprint arXiv:2302.06476, 2023.
- [12] Hendy, A., et al., *How good are gpt models at machine translation? a comprehensive evaluation*. arXiv preprint arXiv:2302.09210, 2023.
- [13] Rudolph, J., S. Tan, and S. Tan, *ChatGPT: Bullshit spewer or the end of traditional assessments in higher education?* Journal of Applied Learning and Teaching, 2023. **6**(1).
- [14] Bang, Y., et al., *A multitask, multilingual, multimodal evaluation of chatgpt on reasoning, hallucination, and interactivity*. arXiv preprint arXiv:2302.04023, 2023.
- [15] Tenhundfeld, N.L., *Two Birds With One Stone: Writing a Paper Entitled "ChatGPT as a Tool for Studying Human-AI Interaction in the Wild" with ChatGPT*.
- [16] Wang, S. and P. Jin, *A Brief Summary of Prompting in Using GPT Models*. 2023.
- [17] George, A.S. and A.H. George, *A review of ChatGPT AI's impact on several business sectors*. Partners Universal International Innovation Journal, 2023. **1**(1): p. 9-23.
- [18] Borji, A., *A categorical archive of ChatGPT failures*. arXiv preprint arXiv:2302.03494, 2023.
- [19] Kalla, D. and N. Smith, *Study and Analysis of Chat GPT and its Impact on Different Fields of Study*. International Journal of Innovative Science and Research Technology, 2023. **8**(3).
- [20] Ray, P.P., *ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope*. Internet of Things and Cyber-Physical Systems, 2023.
- [21] Hariri, W., *Unlocking the Potential of ChatGPT: A Comprehensive Exploration of its Applications, Advantages, Limitations, and Future Directions in Natural Language Processing*. arXiv preprint arXiv:2304.02017, 2023.
- [22] Sebastian, G., *Do ChatGPT and Other AI Chatbots Pose a Cybersecurity Risk?: An Exploratory Study*. International Journal of Security and Privacy in Pervasive Computing (IJSPPC), 2023. **15**(1): p. 1-11.
- [23] Pedreschi, D., et al. *Meaningful explanations of black box AI decision systems*. in *Proceedings of the AAAI conference on artificial intelligence*. 2019.
- [24] Roose, K., *GPT-4 is exciting and scary*. The New York Times, 2023.
- [25] Bubeck, S., et al., *Sparks of artificial general intelligence: Early experiments with gpt-4*. arXiv preprint arXiv:2303.12712, 2023.
- [26] Radford, A., et al., *Language models are unsupervised multitask learners*. OpenAI blog, 2019. **1**(8): p. 9.
- [27] Brown, T., et al., *Language models are few-shot learners*. Advances in neural information processing systems, 2020. **33**: p. 1877-1901.
- [28] Raffel, C., et al., *Exploring the limits of transfer learning with a unified text-to-text transformer*. The Journal of Machine Learning Research, 2020. **21**(1): p. 5485-5551.
- [29] Bolukbasi, T., et al., *Man is to computer programmer as woman is to homemaker? debiasing word embeddings*. Advances in neural information processing systems, 2016. **29**.
- [30] Caliskan, A., J.J. Bryson, and A. Narayanan, *Semantics derived automatically from language corpora contain human-like biases*. Science, 2017. **356**(6334): p. 183-186.
- [31] Blodgett, S.L., et al., *Language (technology) is power: A critical survey of "bias" in nlp*. arXiv preprint arXiv:2005.14050, 2020.
- [32] Solaiman, I., et al., *Release strategies and the social impacts of language models*. arXiv preprint arXiv:1908.09203, 2019.
- [33] Hovy, D. and S. Prabhunoye, *Five sources of bias in natural language processing*. Language and Linguistics Compass, 2021. **15**(8): p. e12432.
- [34] Buolamwini, J. and T. Gebru. *Gender shades: Intersectional accuracy disparities in commercial gender classification*. in *Conference on fairness, accountability and transparency*. 2018. PMLR.
- [35] Bender, E.M. and B. Friedman, *Data statements for natural language processing: Toward mitigating system bias and enabling better science*. Transactions of the Association for Computational Linguistics, 2018. **6**: p. 587-604.
- [36] Kleinberg, J., S. Mullainathan, and M. Raghavan, *Inherent trade-offs in the fair determination of risk scores*. arXiv preprint arXiv:1609.05807, 2016.
- [37] Binns, R. *Fairness in machine learning: Lessons from political philosophy*. in *Conference on fairness, accountability and transparency*. 2018. PMLR.
- [38] Prates, M.O., P.H. Avelar, and L.C. Lamb, *Assessing gender bias in machine translation: a case study with*

- google translate. *Neural Computing and Applications*, 2020. **32**: p. 6363-6381.
- [39] Kirk, H.R., et al., *Bias out-of-the-box: An empirical analysis of intersectional occupational biases in popular generative language models*. *Advances in neural information processing systems*, 2021. **34**: p. 2611-2624.
- [40] Bordia, S. and S.R. Bowman, *Identifying and reducing gender bias in word-level language models*. arXiv preprint arXiv:1904.03035, 2019.
- [41] Johnson, M., et al., *Google's multilingual neural machine translation system: Enabling zero-shot translation*. *Transactions of the Association for Computational Linguistics*, 2017. **5**: p. 339-351.
- [42] Pires, T., E. Schlinger, and D. Garrette, *How multilingual is multilingual BERT?* arXiv preprint arXiv:1906.01502, 2019.
- [43] Dixon, L., et al. *Measuring and mitigating unintended bias in text classification*. in *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*. 2018.
- [44] McGee, R.W., *Is chat gpt biased against conservatives? an empirical study*. *An Empirical Study* (February 15, 2023), 2023.
- [45] Carlini, N., et al. *Extracting Training Data from Large Language Models*. in *USENIX Security Symposium*. 2021.
- [46] Ferrara, E., *The history of digital spam*. *Communications of the ACM*, 2019. **62**(8): p. 82-91.
- [47] Gao, L., et al., *The pile: An 800gb dataset of diverse text for language modeling*. arXiv preprint arXiv:2101.00027, 2020.
- [48] Gururangan, S., et al., *Annotation artifacts in natural language inference data*. arXiv preprint arXiv:1803.02324, 2018.
- [49] Dev, S. and J. Phillips. *Attenuating bias in word vectors*. in *The 22nd International Conference on Artificial Intelligence and Statistics*. 2019. PMLR.
- [50] Wei, J., et al., *Emergent abilities of large language models*. arXiv preprint arXiv:2206.07682, 2022.
- [51] Weidinger, L., et al. *Taxonomy of risks posed by language models*. in *2022 ACM Conference on Fairness, Accountability, and Transparency*. 2022.
- [52] Deng, J. and Y. Lin, *The Benefits and Challenges of ChatGPT: An Overview*. *Frontiers in Computing and Intelligent Systems*, 2022. **2**(2): p. 81-83.
- [53] Brundage, M., et al., *The malicious use of artificial intelligence: Forecasting, prevention, and mitigation*. arXiv preprint arXiv:1802.07228, 2018.
- [54] Ranade, P., et al. *Generating fake cyber threat intelligence using transformer-based models*. in *2021 International Joint Conference on Neural Networks (IJCNN)*. 2021. IEEE.
- [55] Strubell, E., A. Ganesh, and A. McCallum, *Energy and policy considerations for deep learning in NLP*. arXiv preprint arXiv:1906.02243, 2019.
- [56] Adiwardana, D., et al., *Towards a human-like open-domain chatbot*. arXiv preprint arXiv:2001.09977, 2020.
- [57] Howard, J. and S. Ruder, *Universal language model fine-tuning for text classification*. arXiv preprint arXiv:1801.06146, 2018.
- [58] Hassani, H. and E.S. Silva, *The role of ChatGPT in data science: how ai-assisted conversational interfaces are revolutionizing the field*. *Big data and cognitive computing*, 2023. **7**(2): p. 62.
- [59] Rahaman, M., *Can chatgpt be your friend? emergence of entrepreneurial research*. *Emergence of Entrepreneurial Research* (February 18, 2023), 2023.
- [60] Tlili, A., et al., *What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education*. *Smart Learning Environments*, 2023. **10**(1): p. 15.
- [61] Ali, S.R., et al., *Using ChatGPT to write patient clinic letters*. *The Lancet Digital Health*, 2023. **5**(4): p. e179-e181.
- [62] Sallam, M. *ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns*. in *Healthcare*. 2023. MDPI.
- [63] Kung, T.H., et al., *Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models*. *PLoS digital health*, 2023. **2**(2): p. e0000198.
- [64] Khan, R.A., et al., *ChatGPT-Reshaping medical education and clinical management*. *Pakistan Journal of Medical Sciences*, 2023. **39**(2): p. 605.
- [65] Taecharungroj, V., *"What Can ChatGPT Do?" Analyzing Early Reactions to the Innovative AI Chatbot on Twitter*. *Big Data and Cognitive Computing*, 2023. **7**(1): p. 35.
- [66] Tiwary, N., *Netizens, Academicians, and Information Professionals' Opinions About AI With Special Reference To ChatGPT*. arXiv preprint arXiv:2302.07136, 2023.