Navigating the Evolution: A Comprehensive Review of the Technical Progress and Impact of ChatGPT

Jun Lim
Cheriton School of Computer Science
University of Waterloo
jq3lim@uwaterloo.ca

Yingke Wang
Cheriton School of Computer Science
University of Waterloo
y3334wan@uwaterloo.ca

***Abstract*—In the rapidly evolving domain of artificial intelligence, the advent of large language models (LLMs) such as GPT-3 and GPT-4 represents a significant paradigm shift. These models have demonstrated unparalleled linguistic capabilities and led to the creation of tools like ChatGPT, revolutionizing the field of Natural Language Processing (NLP). They have enabled the generation and understanding of text with a sophistication that closely mimics human language, thereby transforming a multitude of AI applications. This study aims to address the lack of scholarly, comprehensive reviews that encapsulate the entire evolution of ChatGPT. Most existing research often limits itself to specific iterations of these models, lacking a review of their history, progress in their technical abilities, and the ethical and societal issues they create. To bridge this gap, this paper offers a systematic and comprehensive review of ChatGPT's evolution, considering its technical, ethical, and societal dimensions. We conduct an in-depth exploration of ChatGPT's journey from its initial stages to its current form as an advanced LLM, addressing critical issues such as the propagation of biases and the potential for misinformation. This review synthesizes studies from a variety of sources to assess the progression and impact of ChatGPT across sectors, including education, healthcare, and business. Through this analysis, the paper aims to deliver a nuanced perspective on ChatGPT's role within the contemporary AI field, emphasizing the need for responsible AI research in light of the social implications caused by these technological breakthroughs.**

***Index Terms*—Artificial Intelligence, Large Language Models (LLMs), Generative Pre-trained Transformers (GPT), Natural Language Processing (NLP), ChatGPT**

I. Introduction

In the past decade, the field of Artificial Intelligence (AI) has experienced a remarkable acceleration, predominantly due to breakthroughs in Natural Language Processing (NLP). The introduction of Generative Pre-trained Transformers (GPT), particularly GPT-3 and GPT-4, has instigated a significant paradigm shift. These models exhibit advanced linguistic capabilities and have led to the development of influential tools like ChatGPT. Characterized by sophisticated algorithms, these models demonstrate an unprecedented proficiency in understanding and generating human-like text, revolutionizing natural language understanding and generation, and opening avenues for diverse applications across multiple sectors [1, 2].

The advent of ChatGPT by OpenAI, a notable advancement in the GPT series, not only showcases the potential technical achievements in AI but also ignites crucial discussions about the ethical, regulatory, and societal implications of such powerful technologies. ChatGPT's facility for enabling seamless human-AI interactions has been applied in various contexts, ranging from simple task automation to complex problem-solving, marking a significant advancement in machine intelligence capabilities [3].

However, the swift development and widespread adoption of these Large Language Models (LLMs) have raised numerous concerns and challenges. Issues ranging from the potential perpetuation of biases to the risks of misinformation have surfaced, fueling debates about the responsible use of AI. As these models increasingly integrate with essential infrastructure and societal frameworks, addressing the resulting issues becomes increasingly imperative [4].

The motivation behind this project stems from the transformative potential of these models across various sectors and the existing gap in comprehensive scholarly reviews that encompass the progress and impact of ChatGPT. From its basic beginnings to the sophisticated mechanisms seen today, understanding its evolution is essential for responsibly harnessing its capabilities.

This study aims to consolidate existing knowledge on the technical evolution of ChatGPT within the framework of Large Language Models (LLMs). It methodically traces ChatGPT's development from its foundational Natural Language Processing (NLP) technologies to the advanced capabilities of the GPT-4 model. Simultaneously, the paper assesses ChatGPT's impact from ethical and societal perspectives. In doing so, we recognize and articulate the significant progress achieved in the field of natural language processing, thereby offering an integrated view of ChatGPT that encompasses both its technological progression and its wider implications [2].

To this end, four research questions (RQs) have been formulated:

1) **RQ1: Evolution of LLMs** - This question investigates the advancements in natural language processing, tracing the progression from earlier models like LSTM [5] to the latest GPT generations. It examines the technical enhancements realized through the adoption of transformer architectures and their contributions to LLMs' development. This inquiry will outline the timeline of improvements in NLP and LLMs, focusing on aspects

such as training architecture, learning methodologies, and the evolution of reasoning abilities.

2) **RQ2: Capabilities of ChatGPT** - Building on RQ1, this question delves into the features that contribute to ChatGPT's technical prowess. We explore how the development of LLMs, including the incorporation of extensive datasets and fine-tuning processes, enhances ChatGPT's capabilities. Specific attention will be paid to how improvements in natural language processing facilitate the handling of complex prompts and broaden the contextual understanding of inputs.
3) **RQ3: Ethical Issues of ChatGPT** - This question conducts a critical analysis of the ethical dilemmas associated with ChatGPT's integration into daily life. As AI systems become more capable and autonomous, concerns around transparency, privacy, and the influence of AI-generated content on public opinion and policy gain prominence. This inquiry will cover a range of ethical issues, including bias, misinformation, and regulatory challenges [6].
4) **RQ4: Societal Impact of ChatGPT** - The final question examines the significant societal dimensions influenced by ChatGPT's introduction. This research explores the wide-ranging impacts of this technology, analyzing its role in transforming educational paradigms, enhancing healthcare services, influencing business and investment strategies, and reshaping the landscape of software engineering [7].

The following parts of this paper is divided into sections. Section II describes background information on our literature reviews. Section III describes our research questions formulated. Section IV elaborates on the research methodology such as its paper search strategy and study selection approach. Then, section V discusses the results obtained after synthesising from the papers obtained, including discussion on the findings of each research question. This is followed by some threats to validity in Section VI and finally a future work and conclusion in section VII and VIII respectively.

II. Background

The field of Artificial Intelligence (AI) has seen exponential growth, particularly with the advent of Natural Language Processing (NLP). NLP has evolved from simple pattern recognition to sophisticated models that understand and generate human-like text. This evolution has been driven by advances in machine learning, especially the development of Large Language Models (LLMs) such as the Generative Pre-trained Transformer (GPT) series by OpenAI. In this section, we will provide an brief background overview on some topics that will be discussed in subsequent sections [8, 9]. This section aims to provide some contextual knowledge on some terminologies and areas.

A. Natural Language Processing

Natural Language Processing (NLP) is a subfield of artificial intelligence (AI) that lies at the intersection of computer science and linguistics, aiming to enable machines to understand and interpret human language as naturally as possible. Initially, the field concentrated on developing rule-based systems, which relied on linguistic theories to parse and generate language structures. Early NLP systems were adept at handling structured tasks like grammar checking and machine translation within limited contexts. However, they struggled with the variability and complexity of natural language, limited by the rigidity of their rule-based frameworks and the sheer diversity of human communication [10].

The advent of statistical NLP in the late 1980s marked a significant paradigm shift. Leveraging the burgeoning availability of digital text, researchers began to train models on large corpora, shifting from hand-crafted rules to data-driven, probabilistic models. This era saw the development of techniques such as part-of-speech tagging and syntactic parsing that took advantage of statistical regularities in text. The application of machine learning to NLP facilitated advancements in language modelling and disambiguation, laying the groundwork for more sophisticated text processing and understanding 3rnfirritchie1990artificial.

B. Machine Learning

Machine Learning (ML) is another field of artificial intelligence (AI) that allows systems to learn and improve from experience without being explicitly programmed. In the context of NLP, ML has been instrumental in advancing the state of the art. Early machine learning approaches in NLP relied on decision trees, support vector machines, and, later, neural networks, each offering incremental improvements in handling language data. The introduction of neural networks, especially recurrent neural networks (RNNs), allowed for the processing of sequential data, opening new avenues for complex language tasks such as speech recognition and machine translation [11].

The breakthrough in ML came with the development of deep learning techniques, which utilize layered neural networks capable of learning high-level features from data. Deep learning has dramatically enhanced the capabilities of NLP systems, enabling them to capture and utilize more abstract patterns within language. As these models grew more sophisticated, they began to outperform traditional ML methods across a variety of NLP tasks, leading to the development of systems that could engage in dialogue, sentiment analysis, and content generation with unprecedented efficacy [12].

C. Large Language Models and GPTs

Large Language Models (LLMs) refer to powerful NLP models that are built using deep learning technologies and trained on massive amounts of text data. LLMs are characterized by their size, complexity and the ability to generate coherent and contextually relevant human-like text. It also represents a significant leap in the machine's capacity to process and generate human language. The architecture commonly used for these models is the transformer architecture, which has proven highly effective in capturing long-range dependencies

in data sequences, making it well-suited for language-related tasks [13].

Generative Pre-trained Transformer (GPT) is a series of transformer-based language models developed by OpenAI. The GPT models are known for generating coherent and contextually relevant text based on a given prompt. The transformer architecture, originally introduced by Vaswani et al. in the paper "Attention is All You Need," [13] forms the backbone of these models. GPT models are trained on vast amounts of text, learning to predict the next word in a sequence, thus generating coherent and contextually relevant language outputs. The first in the series, GPT-1, showcased the potential of transformer architectures to handle long-range dependencies within text, a common challenge in previous models. With each iteration, GPT models have grown in sophistication and scale. GPT-2 expanded the model's capacity and training data, which improved its ability to generate narratives and engage in simple dialogues. GPT-3, however, marked a quantum leap with its 175 billion parameters, enabling it to perform tasks traditionally requiring human-level understanding and creativity [1, 8]. This capacity to generate text that can often pass for human-written has opened new possibilities and challenges in applying AI to real-world problems.

D. ChatGPT

ChatGPT is a state-of-the-art language model that exemplifies the culmination of advancements in NLP and ML, harnessing the power of GPT-3. Designed specifically for conversation, ChatGPT interacts with users in a natural and coherent manner, often indistinguishable from a human. ChatGPT's training includes not just a massive dataset of diverse internet text but also structured dialogues, allowing it to refine its responses based on user interaction. This makes it an invaluable tool for applications ranging from customer service to educational tutoring [1].

The utility of ChatGPT extends beyond mere text generation. It demonstrates an understanding of context, a grasp of nuanced topics, and the ability to maintain the flow of conversation. However, the capabilities of ChatGPT also raise ethical considerations. Its proficiency in mimicking human-like communication brings forth discussions on the responsible use of such technology, particularly concerning biases, the potential for misinformation, and the implications for user privacy and security. ChatGPT, thus, represents not just a technical marvel but also a focal point for broader conversations about the future of AI and its integration into society [14].

III. Research Questions

To guide this review effectively and ensure it captures the essence of ChatGPT's evolution, the following research questions have been formulated. Figure 1 shows the evolutionary timeline of the research questions.

- **RQ1:** What are the improvements made in natural language processing and large language models? **(Technical)**

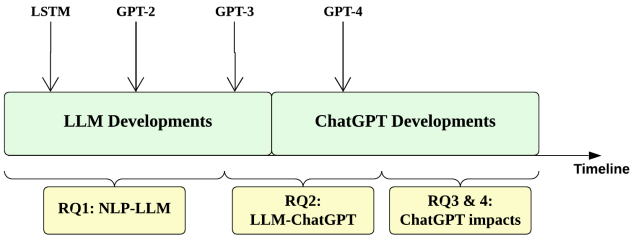

Fig. 1: "Evolutionary" Timeline of Research Questions

- **RQ2:** How have the LLMs advancements contributed to ChatGPT's capabilities? **(Technical)**
- **RQ3:** What are the most common ethical issues of ChatGPT? **(Ethical)**
- **RQ4:** What are the areas most impacted by ChatGPT? **(Societal)**

RQ1 reviews the basic evolution of LLMs. RQ2 extends from RQ1 and investigate "how LLM contributed to ChatGPT development". RQ3 extends from RQ2 and investigates "what are the issues created by the current state of ChatGPT" from an ethical standpoint. RQ4 extends from RQ3 and investigates "despite the issues, what are the impacts created by the current ChatGPT" from a societal standpoint.

IV. Research Methodology

Our research methodology adopts a systematic literature review (SLR) proposed by [15]. Following its guidelines, our methodology included three basic steps, planning the review, conducting the review, and collecting / synthesizing the results, analyzing the results of the reviews.

A. Search Strategy

Figure 3 shows an overall flow of our research methodology, consisting of the flow for identifying the study followed by selection of study.

Through the research questions, we systematically derived search strings for each dimensions covering technical, ethical and societal for papers pertaining to our research. The following search strings are then derived using high-level/broad keywords, as illustrated in Figure 2:

These defined strings are applied onto the automated search of three digital databases: arXiv, ACM Digital Library, and IEEE Xplore. Through this search, **2803 papers are retrieved**.

B. Study Selection

Once the studies that are deemed to be potentially relevant to our studies are retrieved, an assessment of their actual relevance to our research questions is conducted according to the inclusion/exclusion criteria.

Specifically, the following steps were performed, as illustrated in the *Study Selection* of Figure 3:

1) Title, abstract and keywords were inspected to filter out irrelevant papers.
2) Papers that have less than 6 pages are also filtered out.

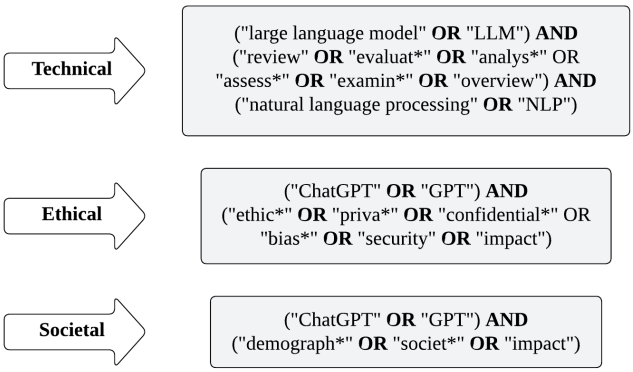

Fig. 2: Query strings for each dimension

3) Full-text of the papers are scanned to obtain papers relevant to the study of reviewing the current state (and evolution) of LLM/ChatGPT.
4) Duplicated papers across multiple databases are filtered out and merged.

Through this four primary steps, guided by the inclusion/exclusion criteria in IV-C, **about 68 papers directly relevant to our research are obtained**.

Of the 68 papers retrieved, about 70% orginates from the database of arXiv, about 20% from ACM Digital Library and 10% from IEEE Xplore, as shown in Figure 4.

C. Inclusion/Exclusion Criteria

Inclusion criteria:

1) The paper claims that an LLM/ChatGPT is used
2) The paper discussed the current state and/or evolution of LLM/ChatGPT
3) The paper discussed technical/ethical/societal aspect of LLM/ChatGPT
4) The paper is accessible with full text

Exclusion criteria:

1) The paper whose number of pages is less than 6
2) Short papers, tutorials, editorials, books or magazines
3) The paper that is published in a workshop or a doctoral symposium
4) The paper is a grey-publication, e.g., technical report or thesis
5) The paper is written in non-English language

V. Results & Discussions

The papers obtained in previous section have been thoroughly reviewed, analyzed and synthesized for its findings to answer the research questions in three dimensions of technical, ethical and societal. The following sections discuss the data synthesized from the papers.

A. *RQ1: What are the improvements made in natural language processing?*

ChatGPT has great power in generating in-context conversational text and logical responses. In this session, we will explore how the field of Natural Language Processing (NLP) evolves through various paradigms shown in Figure 5 from rule-based systems to the recent advent of Large Language Models (LLMs). Both technological advancements and a deeper understanding of language itself drive each stage in the development of NLP. Early systems relied on the manual encoding of linguistic knowledge, but the advent of statistical methods and machine learning has enabled models to learn directly from data.

Nowadays, Large Language Models like GPT [16] and BERT [17] dominate the current state of NLP. They can process and generate language with unprecedented sophistication, often indistinguishable from a human's. However, the rapid progress in NLP has also brought critical issues related to ethics and scalability, prompting the community to focus on creating models that are not only powerful but also responsible and accessible [18].

*1) **Early Beginnings**:* The earliest phase of NLP was characterized by systems that relied on hand-coded rules derived from the principles of computational linguistics. These initial models were designed under the belief that language could be understood through the lens of logic and symbolic presentation. Chomsky's "Syntactic Structures" [19] posited a universal grammar, which attempted to encode grammar rules explicitly.

Researchers such as Allen Newell and Herbert A. Simon also contributed to this early phase with their work on the Logic Theorist and General Problem Solver [20]. This program, encoding human knowledge into machine-readable formats designed to mimic human problem-solving skills, marked the first successful demonstration of AI.

These works laid the groundwork for more complex systems but were not yet dealing extensively with the nuances and complexities of language.

*2) **Rule-Based Systems**:* The development of rule-based systems represented an evolution in the NLP field. These systems flourished in the 1960s and 1970s and utilized extensive lists of rules and dictionaries to parse and interpret language.

These systems saw their use in various applications, from the SHRDLU system, developed by Terry Winograd in 1972, which could understand and respond to natural language commands in a block world [21], to the work of Roger Schank and his conceptual dependency theory for natural language understanding [22]. Despite the sophistication of these systems, they suffered from a lack of flexibility and an inability to cope with the variability of human language. The time and expertise required to craft and update the rules also posed a significant limitation, making these systems impractical for scaling across different languages or domains [10]. Moreover, these systems were constrained by the scalability issue; the more the language complexity grew, the more rules were needed, creating a web of interdependencies that was difficult to manage and expand [23].

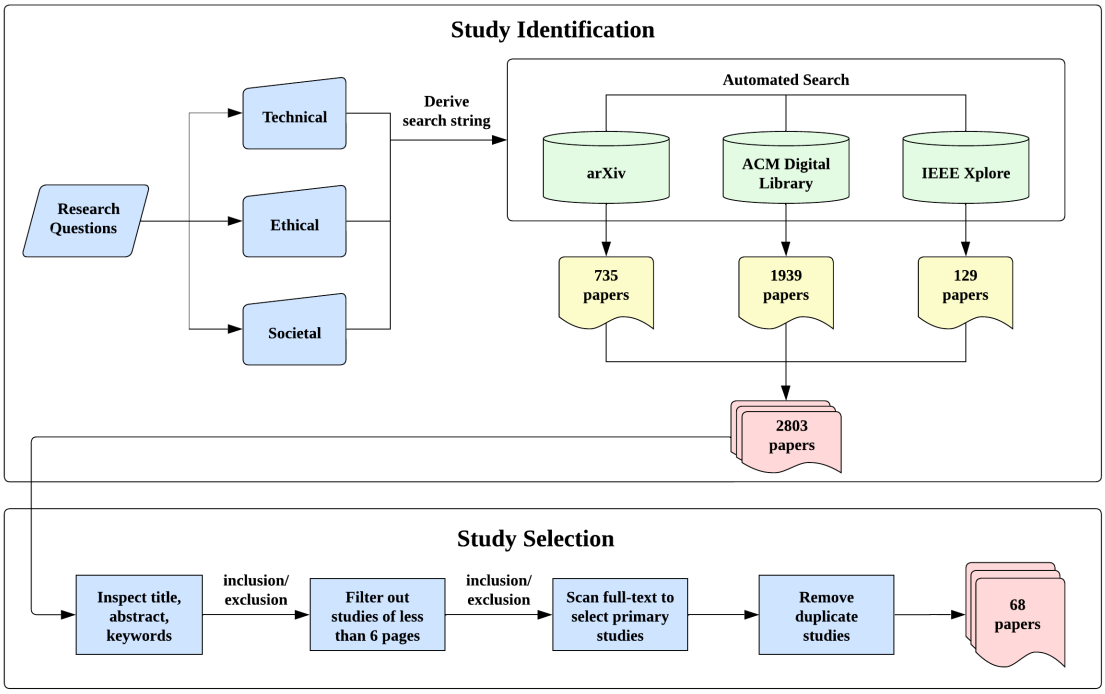

Fig. 3: Research Methodology Flowchart

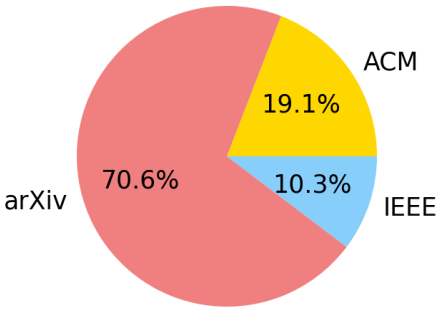

Fig. 4: Distribution of Papers Among Various Databases

3) ***Statistical NLP****:* In the late 1980s and 1990s, there was a major shift towards statical methods in NLP, significantly changing the approach to language analysis. This era was defined by the introduction of machine learning techniques that could automatically learn from data, thus moving away from rigid rule-based systems. A ground-breaking work that contributed to the shift was the introduction of the Hidden Markov Model (HMM) [24]. It provides a probabilistic framework for many NLP tasks.

The use of statistical methods also led to the development of corpus linguistics. In order to efficiently learn useful information from text, large text corpora became valuable resources for linguistic analysis and the training of NLP models. Pann Treebank, one of the most influential corpora developed during this time, greatly improves statistical parsing and part-of-speech tagging [25]. Additionally, using n-gram models for language modelling established the foundation for subsequent machine learning approaches in NLP [26, 27].

4) ***Machine Learning****:* With the advent of the 21st century, the emergence of machine learning as a dominant force in NLP signalled another major transformation. Developing neural networks allowed for modelling text input as high-dimensional continuous vector spaces. Particularly the introduction of word embedding models such as Word2Vec and GloBe, which represent words as dense vectors capturing semantic and syntactic information [28].

Machine learning model architectures also improved during this time. Recurrent neural networks (RNNs) is more adaptive to handling various sequences' length and context in language. These models significantly improved machine translation and other NLP tasks that required an understanding of the sequence and structure of language. However, Standard RNNs suffer from issues like vanishing and exploding gradients, making it hard to learn long-range dependencies in sequences. This led to the development of more advanced RNNs like LSTMs. [29].

5) ***Sequence-to-Sequence Models****:* The mid-2010s saw the rise of sequence-to-sequence (seq2seq) models. These models used a two-part neural network structure with an encoder to process the input sequence and a decoder to generate the output sequence [12]. This design was a significant advancement over previous models because it allowed the entire sequence to be considered, rather than just individual elements, thus capturing long-range dependencies within the text, as

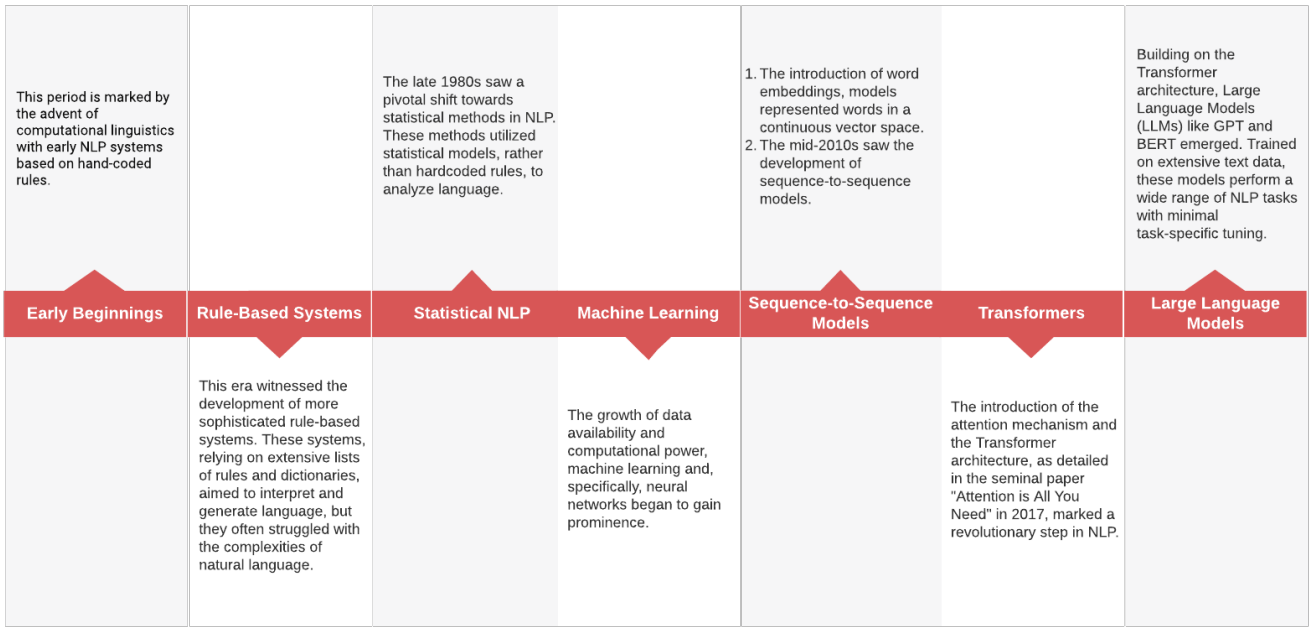

Fig. 5: Natural Language Processing (NLP) Evolution Timeline

illustrated in Figure 6.

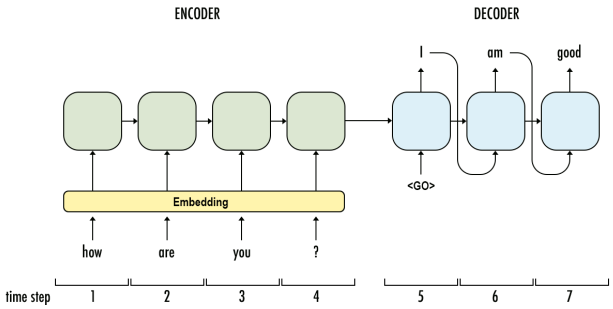

Fig. 6: Sequence-to-Sequence Model Architecture

In detail, The encoder processes the input sequence and compresses the information into a context vector, representing the essence of the input sequence. The decoder then takes this context vector and generates an output sequence, which can be of a different length or nature than the input sequence.

Seq2seq models were particularly impactful in machine translation, where they began to outperform the RNN-based models that had been state-of-the-art for years. The success of seq2seq models in translation was demonstrated by Sutskever et al. in their influential paper in 2014. It showed that a deep LSTM-based seq2seq model could learn to translate with remarkable accuracy [12]. This approach was further refined by introducing attention mechanisms, which allowed the decoder to focus on different parts of the input sequence while generating each word in the output, improving the model's ability to handle long sentences [30].

*6) **Transformers**:* The development of the transformer architecture, introduced in the landmark paper "Attention is All You Need" by Vaswani et al. in 2017, represented a revolutionary step in NLP [13]. Transformers dispensed with the recurrence and convolution of previous models, relying entirely on the attention mechanism to draw global dependencies between input and output. This allowed for more parallelization during training, drastically reducing the time needed to train models while simultaneously improving their performance on tasks like translation, summarization, and text generation.

The transformer model's ability of handle long-range dependencies and its scalability led to the development of models with unprecedented size and capability. Its success has made it the architecture of choice for the latest generation of LLMs such as OpenAI's GPT and Google's BERT [16, 17], setting new standards for what is possible in the field of NLP.

*7) **Large Language Models**:* The emergence of Large Language Models (LLMs) like GPT and BERT marked a new era in NLP. These models, which are characterized by their vast number of parameters and extensive training on large datasets, have significantly advanced across a wide range of NLP tasks. A key to their success is the use of transformer architectures, which allow them to capture the subtle nuances of language by considering the full context of words in their training data. [16, 17]

LLMs are not only capable of performing traditional NLP tasks but have also been adapted for a variety of other applications, demonstrating their versatility. For instance, they have been employed in creating content, from composing music

to drafting legal documents, and in developing sophisticated chatbots and virtual assistants such as ChatGPT [1].

B. RQ2: How have the LLMs advancements contributed to ChatGPT's capabilities?

Diving in to the Large Langurage Models (LLMs) section from RQ1. The Grenerative Pre-trained Transformer (GPT) series by OpenAI have revolutionized natural language processing (NLP), enabling machines to understand and generate human-like text. ChatGPT, a conversational AI developed using these advancements, exemplifies the achievement of LLMs' evolution. This paper aims to explore the progression from GPT-1 to ChatGPT as illustrated in Figure 7 , highlighting how each iteration contributed to refining AI's language abilities and transforming human-AI interactions. The journey from GPT-1 to ChatGPT not only signifies technological advancements but also reflects the changing dynamics of AI in daily life[1, 16].

*1) **GPT-1: The Foundation**:* The first version of the GPT programming language, released to the public in 2018, is built on the Transformer Neural Network architecture designed specifically for natural language processing (NLP) tasks like language modelling and machine translation. GPT-1 underwent pre-training on a vast dataset containing documents, papers, and web content. This training involved the model learning to predict subsequent words in text sequences based on preceding words. GPT-1, through this extensive training, learned to recognize patterns and relationships between words. Further down the stream, GPT-1 could be fine-tuned for specific applications like language translation, text categorization, etc. Despite its relatively modest size of 117 million parameters compared to later GPT versions, GPT-1 demonstrated the value of pre-training on extensive text data for improved language understanding, achieving remarkable performance across a variety of NLP tasks [8, 16].

*2) **GPT-2: Expansion and Improvement**:* Released in 2019, GPT-2 expanded to 1.5 billion parameters, significantly enhancing text generation. The model was pre-trained on a larger corpus of datasets from various sources. Like GPT-1, the model was trained to do the next-word prediction. However, GPT-2 generated more coherent and longer text sequences and demonstrated a greater ability to be utilized in a wider range of domains and tasks. Compared to the previous model, GPT-2 can generate more realistic text aligned with human-written texts. This causes concerns about the potential misuse of the model for generating fake news or scamming. It also brought people's attention to the ethical issues of AI [8, 9, 31].

*3) **GPT-3: Breakthroughs and Capabilities**:* GPT-3, with its massive 175 billion parameters, is significantly larger than its predecessor, GPT-2. The model was trained on an extensive corpus of text data, including web pages, books, and other materials. This training enabled GPT-3 to generate high-quality, coherent, and realistic natural language text. [1] Advancing from the previous model, GPT-3 can do question-answering without requiring task-specific training data. Additionally, GPT-3 incorporates innovative approaches like multi-task learning, enabling the simultaneous performance of multiple tasks, and few-shot learning, which facilitates learning new tasks from minimal examples [32]. These attributes render GPT-3 highly flexible and adaptable for diverse natural language processing applications.

*4) **GPT-3.5: A Pivotal Step Towards ChatGPT**:* GPT-3.5, an enhanced iteration of GPT-3, is pivotal in narrowing the gap between general-purpose language models and specialized conversational AI systems. ChatGPT, the focus of the discussion, is built on the GPT-3.5 framework, a revision of the GPT-3 model initially released by OpenAI in 2020. GPT-3.5 is a more compact model, featuring 6.7 billion parameters compared to the 175 billion parameters of GPT-3 [33–35]. In spite of its reduced parameter count, GPT-3.5 maintains robust performance across a variety of natural language processing tasks. Its capabilities include proficient language understanding, effective text generation, and accurate machine translation [36–38].

*5) **GPT-4: The Latest Advancements**:* The 2023 release of GPT-4 further refined language processing capabilities. This newer version is a substantial multimodal language model capable of processing both text and image inputs to produce text outputs. Although GPT-4 may not match human proficiency in real-world scenarios, it exhibits human-level performance in numerous professional and academic settings. Remarkably, it scored within the top it scored within the top 10% on a simulated bar exam, surpassing the performance of GPT-3.5, which ranked around the bottom 10% [2, 36–38].

*6) **InstructGPT: A Specialized Iteration**:* InstructGPT, developed by OpenAI, represents a significant evolution in the GPT series, specifically designed to enhance instruction-following capabilities in GPT for conversational purposes. This iteration focused on producing responses that are not only accurate but also closely aligned with user intents and instructions, a response to the growing demand for more precise and user-friendly AI interactions. The creation of InstructGPT marked a shift towards useful models in understanding and executing user commands. Such models demonstrated an outstanding ability to perceive and act on user queries, providing context-aware and relevant responses [39, 40]. As InstructGPT can capture human needs and expectations. It sets the stage for creating ChatGPT, which integrates the general language prowess of GPT-3.5 with the instruction-following precision of InstructGPT.

*7) **ChatGPT: Integration of GPT and User Interaction**:* ChatGPT is the synthesizer of the advancements made in the GPT series, integrating the extensive language understanding of GPT-3.5 and GPT-4 with the instruction-following capabili-

Fig. 7: Evolution of GPT Series

ties of InstructGPT. This integration results in a conversational AI that not only understands and generates human-like text but also aligns closely with user intents in a conversation [2, 41]. The development of ChatGPT illustrates a significant evolution in AI capabilities, from basic language processing to engaging in complex, context-aware dialogues. Socially, ChatGPT is also shaping the future of human-AI interaction, setting new standards for AI's role in daily life [42].

*C. **RQ3: What are the most common ethical issues of ChatGPT?***

The ethical papers were systematically and thoroughly analyzed. Each paper was manually read, categorized, and tagged with a set of "ethical issues" keywords. Ethical issues in the context of LLMs or ChatGPT refer to the moral implications and considerations that arise from the development and utilization of these technologies. Some of these ethical issues include, but are not limited to, bias, fairness, privacy, misinformation, transparency, intellectual property, and accountability. After establishing the codes and categories, the frequency of each code is aggregated across all papers, shown in Figure 8

Several recurring themes have emerged from the categorization, with each theme highlighting a distinct ethical concern. The most common issues can be categorized into five broad areas. The following discusses the top five most common ethical issues:

*1) **Bias & Fairness**:* Bias within ChatGPT has emerged as a paramount ethical concern, intricately linked to issues of fairness and equitable treatment. The training of Large Language Models (LLMs) like ChatGPT with existing datasets inevitably leads to the replication of societal biases inherent in these datasets. This critical issue has been highlighted in a substantial body of literature, emphasizing the risks of stereotype perpetuation and challenges in ensuring fair representation by AI systems [7, 43–58].

Numerous studies have identified the consistent negative portrayal of certain groups based on ethnicity, nationality, language, culture, and religion [7, 43, 44, 46, 48, 49, 56], with implications across various sectors including finance, employment, governance, and education. Biases in AI responses become evident when training data is skewed, for instance, predominantly male-focused, leading to the exclusion of women and amplifying gender disparities and discriminatory outcomes [44, 53, 55]. A notable study analyzing ChatGPT's job recommendations revealed biases against different demographic identities, such as recommending lower-paying jobs to individuals of Mexican nationality and secretarial roles predominantly to women [44]. Furthermore, attempts at debiasing through methods like prompt engineering and fine-tuning in job advertisements [47], or incorporating gender disparity data [48], have not shown significant progress in mitigating biases.

The extensive research drawing attention to bias underscores the urgent need for AI tools that are reliable, fair, and inclusive. If these biases remain unaddressed or are used without critical awareness, they risk reinforcing societal stereotypes, marginalizing underrepresented groups, influencing decision-making, and leading to broader sociological impacts [58]. Therefore, despite ChatGPT's impressive performance, it necessitates human oversight to ensure the accuracy and fairness of its responses.

*2) **Misinformation**:* Misinformation within ChatGPT pertains to the generation of false, inaccurate, or misleading information. As ChatGPT formulates responses based on patterns in its training data, it occasionally generates content that, while plausible, is incorrect or deceptive. This issue is particularly alarming in areas where factual accuracy is paramount, such as news dissemination, education, and health advisement [43, 45, 48, 50, 52–57, 59].

A specific study highlighted that AI-generated content is not always accurate and necessitates supplementary background knowledge for verification [59]. Before the advent of ChatGPT, content verification relied on manual methods like consulting credible websites and applying subject matter expertise. However, ChatGPT's capacity to generate a wide array of information might inadvertently lead users into a disinformation trap [51], posing significant risks in critical sectors such as healthcare, aerospace, finance, and nuclear energy if left unchecked [51, 53].

Further research has indicated that ChatGPT's erroneous responses could be exploited by malicious entities to disseminate fake news or propaganda, potentially inflicting harm on individuals and society [52, 53, 55]. For example, the generation of deep fake texts in videos, which are challenging to authenticate, can facilitate the propagation of false narratives [52]. Notably, countries with stringent content censorship, such as Russia, China, North Korea, and Iran, have restricted ChatGPT usage due to concerns over misinformation and the potential for societal destabilization

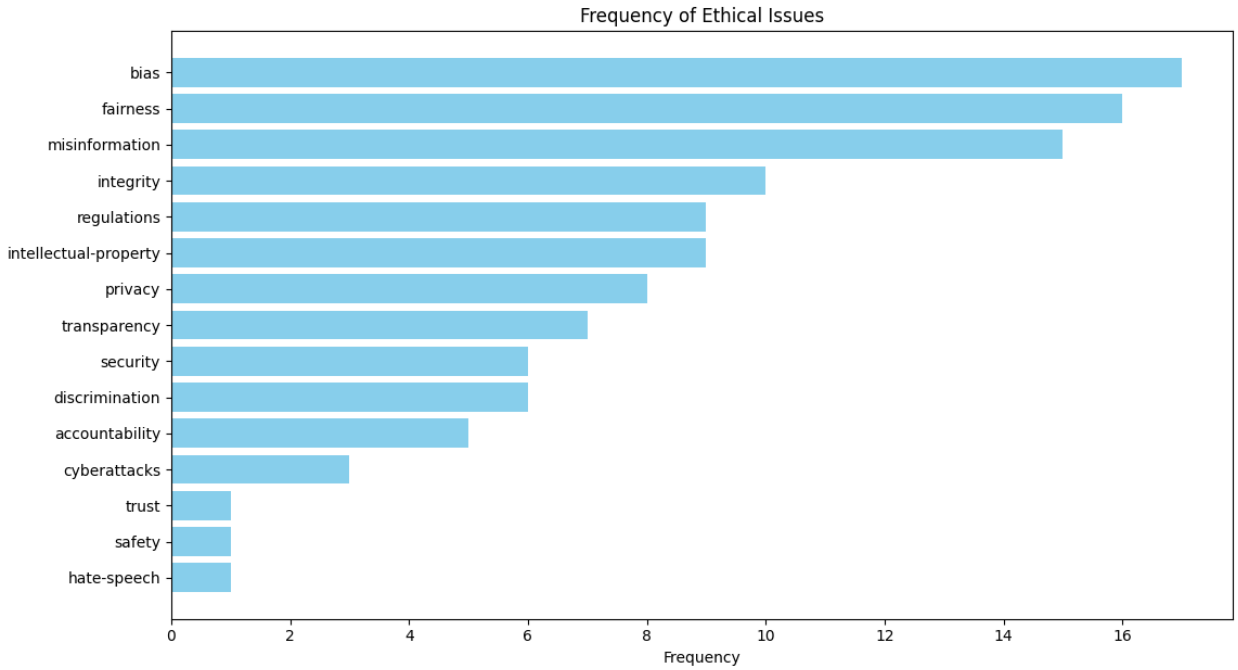

Fig. 8: Frequency of Ethical Issues

[52].

3) ***Integrity****:* The concept of integrity in ChatGPT encompasses the ethical soundness and honesty of AI operations and usage. This concern is particularly prominent in the academic and education sectors, where issues such as unethical practices, illegal activities, dishonesty, academic paper infiltration, and student plagiarism are prevalent [43, 45, 46, 50–55, 59–61].

A case study revealed that ChatGPT is capable of summarizing written texts and enhancing the quality of human-written essays [45]. While tools have been developed to detect plagiarism and integrity violations, their effectiveness is limited, with some unable to distinguish between ChatGPT-generated and human responses and others generating false positives [45, 51]. This undermines the educational goals of fostering critical thinking and independent reasoning, as students may become reliant on ChatGPT for generating their work [59]. Another study demonstrated ChatGPT's ability to deceive a popular plagiarism detection tool, the Measure of Software Similarity (MOSS), especially when responses are minimally edited by humans, leading to the creation of assignments that score highly [60].

In the research community, concerns have been raised about the potential threat to the credibility of researchers who may use AI to augment their work, posing ethical questions regarding the originality of ideas [59]. Moreover, when plagiarized content reaches a corporate scale, it could lead to legal disputes. For example, Amazon has advised its employees against sharing code with ChatGPT, as this might result in ChatGPT storing and training on confidential data, thereby creating ownership and confidentiality issues [52].

Thus, maintaining integrity remains a pressing ethical challenge for ChatGPT. The swift evolution of this technology and the difficulty in regulating these issues in tandem with its development pose significant hurdles in formulating effective solutions.

4) ***Regulations****:* The need for effective regulation is the fifth most cited issue in AI ethics literature [43, 46, 51, 52, 54, 55, 58, 59]. The current regulatory frameworks are struggling to keep pace with rapid advancements in AI technologies, resulting in a governance gap that could lead to ethical transgressions. Legal and ethical frameworks specific to ChatGPT's development, deployment, and usage need to be established and refined to address this emerging challenge.

As AI technologies like ChatGPT become more deeply woven into the fabric of society, the imperative for comprehensive regulations grows. These regulations should encompass privacy, data protection, and the broader impacts of AI on employment, security, and societal welfare [43, 51, 52, 58].

Several solutions have been proposed for AI regulation. These include binding legal instruments, such as copyright laws, administrative measures specific to AI [50, 51, 54, 55], and the establishment of codes of conduct, AI guidelines, and the enforcement of best practices. Yet, implementing these regulations is challenging due to the rapid evolution of AI technologies and the consequent need for laws to adapt swiftly. Additionally, the applicability and effectiveness of AI regulations are largely contingent on the legal context of the country in which they operate. For example, the General Data Protection Regulation (GDPR) in Europe provides robust privacy protections within the European Union (EU), whereas

other jurisdictions, like the California Consumer Privacy Act (CCPA), may not fully address certain privacy and security concerns [50, 51, 55].

Therefore, developing regulations for AI systems such as ChatGPT is an intricate and continuously evolving endeavor. It demands concerted efforts from technologists, ethicists, policymakers, and various stakeholders to establish a unified AI regulatory framework. This framework must integrate a diversity of ethical considerations and cultural perspectives to effectively govern AI [51].

The analysis of these scholarly papers reveals that despite significant advancements in Large Language Models (LLMs), ranging from GPT-2 through GPT-3.5 to GPT-4, ethical considerations continue to be a paramount concern for researchers across various institutions. These concerns emphasize the necessity of ongoing, rigorous evaluations as these models become increasingly integrated into societal applications. This situation underscores the importance of fostering interdisciplinary dialogue and the creation of strong, adaptable frameworks for ethical AI, coupled with effective governance strategies that can keep pace with rapid AI technological advancements. Such measures are crucial to ensure that as LLMs evolve, they align with our collective ethical standards, thereby safeguarding the well-being and rights of individuals and communities.

*D. **RQ4: What are the areas most impacted by ChatGPT?***

Despite numerous authors have highlighted ethical concerns surrounding ChatGPT since its recent public release, yet it has rapidly garnered significant research interest. While the full scope of its influence is still unfolding, it is undeniable that ChatGPT is already transforming several key areas. This section of the research paper delves into the various areas affected by ChatGPT. Employing the same analytical technique as in RQ3, we examine and synthesize relevant literature, with a particular focus on ChatGPT's applications across diverse fields.

The Figure 9 presented illustrates a word cloud, encapsulating the primary areas and applications most significantly influenced by ChatGPT. We further discuss its contributions to various societal sectors and impact.

*1) **Education**:* In an analysis of societal papers, the education sector emerges as the most significantly impacted by ChatGPT, as evidenced by a high volume of citations. ChatGPT has instigated a substantial shift in teaching and learning methodologies within this sector.

For Students: ChatGPT's capacity for providing personalized tutoring and learning support marks a significant advancement. It adapts to varied learning styles, aiding students in comprehending complex topics and enhancing their academic performance. This customized approach is particularly beneficial for diverse learning abilities, potentially reducing the time required for students to complete assignments and master challenging subjects [43, 46, 62–64].

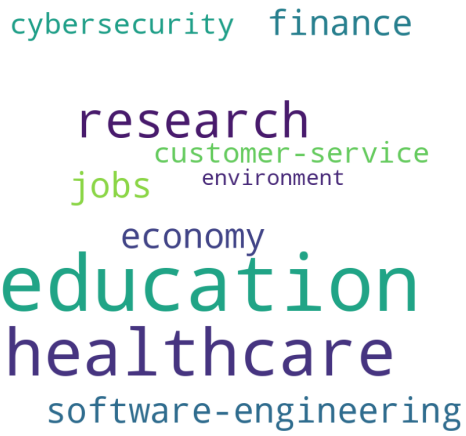

Fig. 9: Word Cloud of Impact Areas

For Educators: ChatGPT serves as an invaluable tool in developing educational materials, grading assignments, and offering feedback. By automating and streamlining numerous educational tasks, it enhances both the efficiency and efficacy of learning environments [43, 54, 64]. Additionally, ChatGPT's role in the creation and evaluation of curricular materials, assessments, and examinations represents a significant time-saving mechanism, enriching the overall teaching experience [43, 64–67].

A primary concern regarding ChatGPT in education centers on the potential diminution of creativity and critical thinking skills. There is a risk of students becoming overly reliant on AI for assignment completion, which could lead to diminished engagement and academic integrity issues [59, 63, 68–70]. The challenge of plagiarism detection in AI-generated student work is notable, as tools like MOSS (Measure of Software Similarity) struggle to identify content produced by GPT, raising questions about the integrity of student work and the efficacy of existing plagiarism detection methods [60, 62]. Additionally, the dissemination of inaccurate information and the risk of data breaches are significant concerns in the sensitive context of academia, necessitating vigilant monitoring and verification of AI-generated content [45, 70].

Despite these challenges, research indicates that AI technologies like ChatGPT can facilitate equitable educational access by providing resources and support to a broader spectrum of students, including those who may lack traditional educational opportunities [61]. In conclusion, ChatGPT's impact on education is multifaceted and complex, presenting both opportunities and challenges, and the education sector must therefore carefully balance the benefits and risks associated with AI integration.

*2) **Healthcare**:* The integration of ChatGPT into the healthcare sector represents a significant area of impact. Since its inception, ChatGPT has found diverse applications in healthcare, ranging from telemedicine and medical research to education, data analytics, and serving as a 24/7 health assistant [43, 63, 69, 71].

In healthcare services, numerous medical facilities have embraced AI transformation by:

1) Offering personalized healthcare and medication recommendations,
2) Utilizing virtual health assistants for responding to medical inquiries,
3) Innovating in clinical process development and service provision,
4) Assisting in medical diagnostics,
5) Engaging in predictive analytics [63, 65, 71, 72].

In clinical processes, a study focusing on clinical workflow optimization in medical imaging revealed that ChatGPT can efficiently and accurately analyze data. This capability enables it to reduce time spent on tasks like interpreting radiology scans and results, and in generating automated reports [72]. Other studies have observed enhancements in efficiency in areas such as patient triage, symptom analysis, and medical diagnosis [64].

In medical research and education, healthcare researchers utilize it for tasks like evaluating research outcomes, summarizing research reports, and providing detailed answers on specific study topics [65, 72]. Medical students have found it particularly beneficial in assisting with clinical trials and in augmenting their medical training and expertise [64].

Despite its impact, several studies have raised significant ethical concerns related to the reliability, safety, and factuality of the content generated by ChatGPT [63, 65, 73]. There is a risk of disseminating false information regarding medications, medical conditions, or other medically-related information, potentially misleading patients and leading to adverse outcomes [63, 65]. Additionally, the deployment of ChatGPT in healthcare is fraught with regulatory issues, including concerns about data transparency, protection, and potential biases in medical responses [64, 69]. Addressing these critical issues is essential before considering the full-scale implementation of ChatGPT in medical services.

3) ***Research****:* ChatGPT's proficiency in language translation, summarization, and data analysis renders it an indispensable tool in interdisciplinary research [45, 63, 65, 72, 74]. It facilitates collaboration across diverse fields by enabling researchers to effectively communicate and analyze complex information. For instance, in joint projects involving linguists and data scientists, ChatGPT can translate specialized jargon into understandable language and analyze data patterns, promoting an integrated research approach [45, 63].

The conventional literature review process, which is often exhaustive and time-intensive, is streamlined by ChatGPT. It efficiently identifies relevant literature and succinctly summarizes key findings. This approach not only expedites research but also broadens the scope of sources reviewed, resulting in more thorough literature reviews and well-informed research conclusions [45]. Additionally, ChatGPT's ability to analyze extensive datasets can reveal patterns and insights possibly overlooked by human researchers, proving particularly valuable in data-rich fields like genomics or climate research [45, 63].

In sectors such as healthcare and AI, ChatGPT offers new methods for conceptualizing and designing research. It can propose innovative research questions and methods, potentially leading to groundbreaking discoveries and fostering creativity in problem-solving [72, 74]. For example, a study highlighted how the integration of ChatGPT and AI technologies has expedited the publication of AI research, indicative of an upward trend in research scale [74].

However, there is a concern regarding potential overreliance on ChatGPT for intellectually demanding tasks, which could impede the development of critical analytical skills essential in scientific inquiry [64]. Notably, ChatGPT has been utilized as a co-author in scientific articles due to its coherent content generation capabilities [45]. This practice has ignited a debate in the academic community, with some journals imposing restrictions on ChatGPT's use as a co-author, questioning the originality of AI-generated research content. This controversy underscores the necessity for establishing explicit guidelines for AI use in research writing.

While ChatGPT offers significant benefits, it is imperative to acknowledge its limitations, particularly concerning the accuracy of its responses. Sole reliance on ChatGPT without thorough evaluation and validation of its outputs might compromise research quality [63].

4) ***Business & Investments****:* The business industry is multifaceted, encompassing various fields such as finance, investments, marketing, and e-commerce. In these domains, ChatGPT demonstrates its capability by analyzing extensive historical data to forecast market trends, optimizing inventory management to reduce costs, and assessing supply chain risks [69]. Such analytical proficiency aids in making informed decisions, thereby increasing efficiency and reducing business expenses. Specifically, ChatGPT's ability to predict financial market trends is highly advantageous for investors, financial analysts, and portfolio managers, contributing to the growth of a country's financial markets.

In the realm of financial investment, ChatGPT's impact is transformative. A study examining the market efficiency of an AI crypto index, utilizing the GPT model of ChatGPT, indicated a significant increase in profit returns [75] This showcases ChatGPT's capability in performing complex predictive analyses. Additionally, an increasing trend has been observed where financial services are integrating virtual financial advisors, or robo-advisors, equipped with sophisticated AI-based prediction algorithms [63]. This integration facilitates the generation of personalized investment recommendations tailored to individual financial goals and risk tolerance. Furthermore, in financial research, ChatGPT plays a pivotal role by aiding researchers in analyzing vast quantities of cryptocurrency research, thereby contributing to the development of more robust financial strategies [63].

However, the extensive use of ChatGPT in investments is not without its concerns. Issues such as potential algorithmic

biases, cybersecurity risks, and an overreliance on AI could negatively affect business operations if not carefully managed [69]. For example, malfunctions in AI servers could disrupt business activities and lead to significant financial losses for organizations. Therefore, it is imperative to maintain a balanced approach, integrating human judgment with AI technology in business and investment strategies.

*5) **Software Engineering**:* In the realm of software engineering, ChatGPT has markedly transformed coding and development processes, including code generation, debugging, testing, documentation, and review [46, 64]. Traditional software development, characterized by intensive time investment and manual coding, is evolving with ChatGPT's intervention. Developers can now rapidly generate code from high-level descriptions, expediting the initial development stages. This accelerates prototyping, enabling quicker testing and iteration of ideas, fostering more agile development cycles.

ChatGPT's significant contribution to software development is evident, particularly through platforms like CodeStarter. One study highlights CodeStarter's ability to write scripts for various frameworks based on a simple web app description, thus saving considerable developer time and effort [46]. A notable portion of the code produced by such tools is of high quality, underscoring ChatGPT's efficiency in coding. Additionally, ChatGPT streamlines debugging, enhancing software quality and reliability. However, its proficiency varies across domains; for instance, ChatGPT correctly answered only 55.6% of textbook questions on software testing, revealing some limitations [45].

In documentation, ChatGPT excels at creating technical documentation for software projects and providing programming guidance. It particularly benefits developers with limited experience, allowing them to focus on more complex tasks by handling routine activities like syntax checking and code assistance [68]. Consequently, developers can devote more time to the intricate facets of software development.

Nevertheless, ChatGPT's role in development also introduces intellectual property (IP) and ethical considerations. Developers utilizing ChatGPT outputs must navigate potential copyright issues [68], ensuring compliance with IP laws and appropriate attribution. This underscores the need for a mindful approach to AI tool utilization, promoting ethical AI practices in software development.

While ChatGPT offers promising advancements in software engineering, its increasing influence also presents challenges. These include ensuring the precision of AI-generated code, addressing ethical and IP concerns, and balancing automation with human oversight [54].

VI. Threats to Validity

A. Rapid Pace of Development

The rapid pace of development of ChatGPT and related technologies significantly challenges the validity of this study. This lightning-fast evolution means that the research landscape is constantly changing, with new findings and advancements emerging regularly. Consequently, this review, which relies on literature from 2020 to 2023, may not fully encapsulate the latest developments in the field, such as its emerging trends. As such, the findings here might become outdated quickly as new research and breakthroughs in LLMs continuously emerge. Therefore, while this review aims to be comprehensive, it inherently represents a snapshot that is at risk of becoming quickly outdated due to the field's dynamic nature.

B. Study Selection Bias

Our methodology for selecting studies for this literature review is primarily based on keyword searches, which introduces potential biases because the selection process is heavily dependent on the choice of keywords, which might not capture all the relevant literature in the field. Unlike other systematic literature review processes, we did not adopt a diversified search strategy that included a combination of manual searches (from conferences and journals), automated searches, and snowballing techniques, which increased the risk of missing out on relevant literature. While our methodology adopted a broad-search strategy to cover as many relevant keywords as possible in each dimension, its inherent limitations might still result in a skewed representation of the research, which could affect the validity of the research.

C. Inclusion of Preprints

The inclusion of preprints from sources such as arXiv, which are not yet peer-reviewed, poses a challenge to the reliability of the findings. These documents offer insights into the latest research but may contain preliminary findings that have not undergone rigorous academic scrutiny. While the number of preprints is relatively small compared to peer-reviewed publications, their inclusion necessitates careful interpretation of the study's findings, as they could introduce elements of uncertainty and affect the overall quality of the conclusions drawn.

D. Manual Study and Categorization

The manual process of selecting and categorizing studies for the review, as described in Section IV, carries the risk of subjective biases as the selection and interpretation of studies could be influenced by the reviewers' judgments. To mitigate this risk, a double-review process was employed, where each study was initially reviewed and then subjected to a secondary assessment by a different reviewer. This approach aimed to enhance the reliability of the report by minimizing the potential for individual biases. However, despite these precautions, the possibility of judgment biases influencing the study interpretation cannot be entirely eliminated, which remains a limitation that underscores the need for cautious interpretation of the review findings.

VII. Future Work

This comprehensive study of ChatGPT's evolution and its current state in various sectors lays the groundwork for potential future research avenues. Firstly, as AI models continue

to grow in sophistication, a focused study on the development of more robust frameworks for detecting and mitigating bias in language models is vital. For instance, this could include the creation of diverse datasets that better represent a broader population and the continuous evaluation of model outputs for fairness and inclusivity [6].

Secondly, the intersection of ChatGPT with critical sectors like healthcare and education could be extended for domain-specific studies. These could include assessing the long-term impact of AI assistance on professional roles, patient outcomes, student learning, and the possibility of job displacement and creation. Investigating these impacts will provide deeper insights into the sustainable integration of AI technologies into human-centric fields [36, 38, 42].

Thirdly, the ethical and regulatory challenges highlighted a need for interdisciplinary research to develop comprehensive guidelines and standards. Future work could explore how policy-making can evolve in pace with AI advancements, ensuring that governance keeps pace with innovation. Furthermore, the potential for spreading misinformation through AI-generated content also suggested a possible future research agenda. For instance, we could investigate and explore methods for verifying the authenticity of AI information and tools for educating the public on discerning such content [42].

Lastly, given the rapid pace of development in AI, there is also a need to have a continuous re-evaluation of the state-of-the-art. This includes continuous studies to track the evolution of ChatGPT models and their societal impact in order to provide up-to-date, valuable feedback for AI developers, users, and policymakers [7].

VIII. Conclusion

In summarizing the technical evolution and societal impact of ChatGPT, this paper highlights the remarkable advancements in Natural Language Processing achieved through Generative Pre-trained Transformers, particularly GPT-4. While ChatGPT's integration across various sectors underscores its transformative potential, it also brings to the forefront critical ethical considerations like bias, privacy, and misinformation. This study emphasizes the need for ongoing, responsible development of AI technologies, balancing innovation with mindful regulation. As ChatGPT continues to advance, it is essential to critically navigate its trajectory, ensuring that its benefits are maximized while mitigating potential adverse societal impacts [33].

References

[1] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, ..., and D. Amodei, "Language models are few-shot learners," in *Advances in Neural Information Processing Systems*, 2020.

[2] R. OpenAI, "Gpt-4 technical report. arxiv 2303.08774," *View in Article*, vol. 2, p. 3, 2023.

[3] F. Dennstädt, J. Hastings, P. M. Putora, E. Vu, G. Fischer, K. Süveg, M. Glatzer, E. Riggenbach, H.-L. Hà, and N. Cihoric, "Exploring the capabilities of large language models such as chatgpt in radiation oncology," *Advances in Radiation Oncology*, p. 101400, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2452109423002282

[4] K. Crawford and R. Calo, "There is a blind spot in ai research," *Nature*, vol. 538, pp. 311–313, 10 2016.

[5] R. C. Staudemeyer and E. R. Morris, "Understanding lstm – a tutorial into long short-term memory recurrent neural networks," 2019.

[6] I. O. Gallegos, R. A. Rossi, J. Barrow, M. M. Tanjim, S. Kim, F. Dernoncourt, T. Yu, R. Zhang, and N. K. Ahmed, "Bias and fairness in large language models: A survey," 2023.

[7] A. Tamkin, M. Brundage, J. Clark, and D. Ganguli, "Understanding the capabilities, limitations, and societal impact of large language models," 2021.

[8] X. Zheng, C. Zhang, and P. C. Woodland, "Adapting gpt, gpt-2 and bert language models for speech recognition," in *2021 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, 2021, pp. 162–168.

[9] Y. Qu, P. Liu, W. Song, L. Liu, and M. Cheng, "A text generation and prediction system: pre-training on new corpora using bert and gpt-2," in *2020 IEEE 10th International Conference on Electronics Information and Emergency Communication (ICEIEC)*. IEEE, July 2020, pp. 323–326.

[10] G. Gazdar and C. Mellish, *Natural Language Processing in Lisp: An Introduction to Computational Linguistics*. Addison-Wesley, 1989.

[11] D. Operationnelle, Y. Bengio, R. Ducharme, P. Vincent, and C. Mathematiques, "A neural probabilistic language model," 10 2001.

[12] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Advances in Neural Information Processing Systems*, 2014.

[13] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, 2017, pp. 5998–6008.

[14] H. Chen, F. Jiao, X. Li, C. Qin, M. Ravaut, R. Zhao, C. Xiong, and S. Joty, "Chatgpt's one-year anniversary: Are open-source large language models catching up?" 2023.

[15] S. Keele *et al.*, "Guidelines for performing systematic literature reviews in software engineering," 2007.

[16] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving language understanding by generative pre-training," 2018.

[17] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[18] M. Whittaker *et al.*, *AI now report 2018*, 2018.

[19] N. Chomsky, *Syntactic Structures*. The Hague/Paris: Mouton, 1957.

[20] A. Newell and H. A. Simon, "The logic theorist," RAND Corporation, Tech. Rep., 1956.

[21] T. Winograd, "Understanding natural language," *Cognitive Psychology*, 1972.

[22] R. C. Schank, "Conceptual dependency: A theory of natural language understanding," *Cognitive Psychology*, 1972.

[23] G. Ritchie and G. Russell, "Artificial intelligence: A historical perspective," *AI Society*, vol. 4, 1990.

[24] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, 1989.

[25] M. P. Marcus, B. Santorini, and M. A. Marcinkiewicz, "Building a large annotated corpus of english: The penn treebank," *Computational Linguistics*, 1993.

[26] F. Jelinek, L. R. Bahl, and R. L. Mercer, "A maximum likelihood approach to continuous speech recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1980.

[27] C. D. Manning and H. Schütze, *Foundations of Statistical Natural Language Processing*. MIT Press, 1999.

[28] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," in *Proceedings of Workshop at ICLR*, 2013.

[29] A. Sherstinsky, "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network," *CoRR*, vol. abs/1808.03314, 2018. [Online]. Available: http://arxiv.org/abs/1808.03314

[30] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.

[31] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, "Language models are unsupervised multitask learners," OpenAI Blog, 2019.

[32] J. Hewett and M. Leeke, "Developing a gpt-3-based automated victim for advance fee fraud disruption," in *2022 IEEE 27th Pacific Rim*

International Symposium on Dependable Computing (PRDC). IEEE, November 2022, pp. 205–211.
[33] A. Borji, "A categorical archive of chatgpt failures," *arXiv preprint arXiv:2302.03494*, 2023.
[34] H. Alkaissi and S. I. McFarlane, "Artificial hallucinations in chatgpt: implications in scientific writing," *Cureus*, vol. 15, no. 2, 2023.
[35] S. Frieder, L. Pinchetti, R. R. Griffiths, T. Salvatori, T. Lukasiewicz, P. C. Petersen, A. Chevalier, and J. Berner, "Mathematical capabilities of chatgpt," *arXiv preprint arXiv:2301.13867*, 2023.
[36] D. Baidoo-Anu and L. Owusu Ansah, "Education in the era of generative artificial intelligence (ai): Understanding the potential benefits of chatgpt in promoting teaching and learning," SSRN 4337484, 2023.
[37] D. R. Cotton, P. A. Cotton, and J. R. Shipway, "Chatting and cheating: Ensuring academic integrity in the era of chatgpt," *Innovations in Education and Teaching International*, pp. 1–12, 2023.
[38] A. Howard, W. Hope, and A. Gerada, "Chatgpt and antimicrobial advice: the end of the consulting infection doctor?" *Lancet Infectious Diseases*, 2023.
[39] B. Bhavya, J. Xiong, and C. Zhai, "Analogy generation by prompting large language models: A case study of instructgpt," *arXiv preprint arXiv:2210.04186*, 2022.
[40] A. Chan, "Gpt-3 and instructgpt: Technological dystopianism, utopianism, and "contextual" perspectives in ai ethics and industry," *AI and Ethics*, pp. 1–12, 2022.
[41] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. L. Wainwright, P. Mishkin, ..., and M. Chen, "Training language models to follow instructions with human feedback," 2022.
[42] M. Abdullah, A. Madain, and Y. Jararweh, "Chatgpt: fundamentals, applications and social impacts," in *2022 Ninth International Conference on Social Networks Analysis, Management and Security (SNAMS)*. IEEE, November 2022, pp. 1–8.
[43] M. T. Baldassarre, D. Caivano, B. Fernandez Nieto, D. Gigante, and A. Ragone, "The social impact of generative ai: An analysis on chatgpt," in *Proceedings of the 2023 ACM Conference on Information Technology for Social Good*, ser. GoodIT '23. New York, NY, USA: Association for Computing Machinery, 2023, p. 363–373.
[44] A. Salinas, P. Shah, Y. Huang, R. McCormack, and F. Morstatter, "The unequal opportunities of large language models: Examining demographic biases in job recommendations by chatgpt and llama," in *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, ser. EAAMO '23. New York, NY, USA: Association for Computing Machinery, 2023.
[45] S. Shahriar and K. Hayawi, "Let's have a chat! a conversation with chatgpt: Technology, applications, and limitations," *Artificial Intelligence and Applications*, 2023.
[46] M. Zong and B. Krishnamachari, "a survey on gpt-3," 2022.
[47] C. Borchers, D. S. Gala, B. Gilburt, E. Oravkin, W. Bounsi, Y. M. Asano, and H. R. Kirk, "Looking for a handsome carpenter! debiasing gpt-3 job advertisements," 2022.
[48] U. Gupta, J. Dhamala, V. Kumar, A. Verma, Y. Pruksachatkun, S. Krishna, R. Gupta, K.-W. Chang, G. V. Steeg, and A. Galstyan, "Mitigating gender bias in distilled language models via counterfactual role reversal," 2022.
[49] L. Magee, L. Ghahremanlou, K. Soldatic, and S. Robertson, "Intersectional bias in causal language models," 2021.
[50] I. Ullah, N. Hassan, S. S. Gill, B. Suleiman, T. A. Ahanger, Z. Shah, J. Qadir, and S. S. Kanhere, "Privacy preserving large language models: Chatgpt case study based vision and framework," 2023.
[51] X. Wu, R. Duan, and J. Ni, "Unveiling security, privacy, and ethical concerns of chatgpt," 2023.
[52] A. Qammar, H. Wang, J. Ding, A. Naouri, M. Daneshmand, and H. Ning, "Chatbots to chatgpt in a cybersecurity space: Evolution, vulnerabilities, attacks, challenges, and future recommendations," 2023.
[53] M. A. Akbar, A. A. Khan, and P. Liang, "Ethical aspects of chatgpt in software engineering research," 2023.
[54] L. Yan, L. Sha, L. Zhao, Y. Li, R. Martinez-Maldonado, G. Chen, X. Li, Y. Jin, and D. Gašević, "Practical and ethical challenges of large language models in education: A systematic scoping review," *British Journal of Educational Technology*, Aug. 2023.
[55] Y. Wang, Y. Pan, M. Yan, Z. Su, and T. H. Luan, "A survey on chatgpt: Ai–generated contents, challenges, and solutions," *IEEE Open Journal of the Computer Society*, vol. 4, pp. 280–302, 2023.
[56] T. Y. Zhuo, Y. Huang, C. Chen, and Z. Xing, "Red teaming chatgpt via jailbreaking: Bias, robustness, reliability and toxicity," 2023.
[57] C. Wald and L. Pfahler, "Exposing bias in online communities through large-scale language models," 2023.
[58] V. Thakur, "Unveiling gender bias in terms of profession across llms: Analyzing and addressing sociological implications," 2023.
[59] A. R. Vargas-Murillo, I. N. M. d. l. A. Pari-Bedoya, and F. d. J. Guevara-Soto, "The ethics of ai assisted learning: A systematic literature review on the impacts of chatgpt usage in education," in *Proceedings of the 2023 8th International Conference on Distance Education and Learning*, ser. ICDEL '23. New York, NY, USA: Association for Computing Machinery, 2023, p. 8–13.
[60] S. Biderman and E. Raff, "Fooling moss detection with pretrained language models," in *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, ser. CIKM '22. New York, NY, USA: Association for Computing Machinery, 2022, p. 2933–2943.
[61] F. Draxler, D. Buschek, M. Tavast, P. Hämäläinen, A. Schmidt, J. Kulshrestha, and R. Welsch, "Gender, age, and technology education influence the adoption and appropriation of llms," 2023.
[62] K. Malinka, M. Perešíni, A. Firc, O. Hujnák, and F. Janus, "On the educational impact of chatgpt: Is artificial intelligence ready to obtain a university degree?" in *Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 1*, ser. ITiCSE 2023. ACM, Jun. 2023.
[63] W. Hariri, "Unlocking the potential of chatgpt: A comprehensive exploration of its applications, advantages, limitations, and future directions in natural language processing," 2023.
[64] M. Fraiwan and N. Khasawneh, "A review of chatgpt applications in education, marketing, software engineering, and healthcare: Benefits, drawbacks, and research directions," 2023.
[65] I. S. Gabashvili, "The impact and applications of chatgpt: a systematic review of literature reviews," 2023.
[66] N. Y. Motlagh, M. Khajavi, A. Sharifi, and M. Ahmadi, "The impact of artificial intelligence on the evolution of digital education: A comparative study of openai text generation tools including chatgpt, bing chat, bard, and ernie," 2023.
[67] A. M. Abdelfattah, N. A. Ali, M. A. Elaziz, and H. H. Ammar, "Roadmap for software engineering education using chatgpt," in *2023 International Conference on Artificial Intelligence Science and Applications in Industry and Society (CAISAIS)*, 2023, pp. 1–6.
[68] C.-N. Anagnostopoulos, "Chatgpt impacts in programming education: A recent literature overview that debates chatgpt responses," 2023.
[69] A. Bahrini, M. Khamoshifar, H. Abbasimehr, R. J. Riggs, M. Esmaeili, R. M. Majdabadkohne, and M. Pasehvar, "Chatgpt: Applications, opportunities, and threats," 2023.
[70] R. H. Mogavi, C. Deng, J. J. Kim, P. Zhou, Y. D. Kwon, A. H. S. Metwally, A. Tlili, S. Bassanelli, A. Bucchiarone, S. Gujar, L. E. Nacke, and P. Hui, "Exploring user perspectives on chatgpt: Applications, perceptions, and implications for ai-integrated education," 2023.
[71] G. Zuccon and B. Koopman, "Dr chatgpt, tell me what i want to hear: How prompt knowledge impacts health answer correctness," 2023.
[72] J. Yang, H. B. Li, and D. Wei, "The impact of chatgpt and llms on medical imaging stakeholders: Perspectives and use cases," *Meta-Radiology*, vol. 1, no. 1, p. 100007, Jun. 2023.
[73] Z. Yan, K. Zhang, R. Zhou, L. He, X. Li, and L. Sun, "Multimodal chatgpt for medical applications: an experimental study of gpt-4v," 2023.
[74] F. Joublin, A. Ceravola, J. Deigmoeller, M. Gienger, M. Franzius, and J. Eggert, "A glimpse in chatgpt capabilities and its impact for ai research," 2023.
[75] A. Zamfiroiu, D. Vasile, and D. Savu, "Chatgpt–a systematic review of published research papers," *Informatica Economica*, vol. 27, no. 1, pp. 5–16, 2023.