



## Original Article

## The limitations and constraints of ChatGPT: an annotated bibliography for researchers and professionals

Shaoyi He<sup>1</sup> 

### Abstract

This annotated bibliography delves into the intricate web of limitations and constraints that surround ChatGPT, a pioneering natural language processing AI model. As AI technologies continue to redefine human-computer interaction, understanding the boundaries and potential pitfalls of these systems becomes paramount. This comprehensive resource compiles a diverse array of sources, including academic studies, reports, and analyses, published between December 2022 and August 2023 and dedicated to unraveling specific facets of ChatGPT's limitations and constraints. Annotated entries not only provide succinct summaries of these limitations but also offer critical assessments, shedding light on their significance, credibility, and real-world implications. The annotated bibliography aims to be an essential guide for researchers, developers, policymakers, and users, navigating the complexities of ChatGPT's constraints while fostering responsible AI development, informed decision-making, and ethical AI usage. In an era where AI plays an increasingly integral role in our lives, understanding the limitations and constraints of such technologies is not merely an academic exercise but a crucial step toward harnessing their potential for the betterment of society.

**Key Words:** annotated bibliography; ChatGPT; limitations; constraints; researchers; professionals

---


**Affiliation Info:** <sup>1</sup> Jack H. Brown College of Business and Public Administration, California State University, San Bernardino, USA

**Article Info:** 20 November 2023 / Revised: 6 December 2023 / Accepted: 10 December 2023 / Published Online: 12 December 2023. [www.naturalisscientias.com](http://www.naturalisscientias.com)

**Authors' Contact Info:** He, S: [she@csusb.edu](mailto:she@csusb.edu)

**Citation:** He, S. 2024. The limitations and constraints of ChatGPT: an annotated bibliography for researchers and professionals. *Naturalis Scientias*, 1(1): 106-130. DOI: <https://doi.org/10.62252/NSS.2024.1008>.

**Copyright** © 2024 by the author. Published by Naturalis Scientias. This is an open access article under the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License. (<https://creativecommons.org/licenses/by-nc/4.0/>).

**Corresponding Author** : He, S; Email: [she@csusb.edu](mailto:she@csusb.edu)

## 1 Introduction

ChatGPT, an acronym for Chat Generative Pre-trained Transformer, represents a substantial language model-driven chatbot that was developed by OpenAI and unveiled on November 30, 2022. It empowers users to shape and guide conversations according to their preferences, allowing control over factors such as conversation length, format, style, level of detail, and language employed. In the rapidly evolving landscape of artificial intelligence, ChatGPT stands as a testament to the transformative power of natural language processing models. Its capacity to generate coherent and contextually relevant text responses has ushered in a new era of human-computer interaction, touching countless aspects of our lives, from education and customer service to content creation and beyond. Yet, amid the euphoria surrounding the capabilities of ChatGPT, it is imperative not to overlook its inherent limitations and constraints.

This annotated bibliography represents an earnest endeavor to dive deep into the multifaceted boundaries and constraints that ChatGPT encounters in its applications. While the AI community has celebrated its achievements, it has also engaged in rigorous scrutiny, producing an array of insightful studies, reports, and analyses. This bibliography serves as a curated compendium of these invaluable sources, offering a comprehensive exploration of ChatGPT's limitations and constraints.

The curated annotated entries within this bibliography span a spectrum of perspectives and research methodologies, each contributing to a nuanced understanding of ChatGPT's constraints. These sources explore diverse facets of the technology's limitations and constraints, ranging from its contextual comprehension challenges and susceptibility to biases to its difficulties in handling sensitive data and the ethical implications of its usage.

Our goal here is twofold: firstly, to provide a consolidated repository of these sources, enabling researchers, developers, policymakers, and users to navigate the complex terrain of ChatGPT's constraints with ease, and secondly, to offer critical assessments of these sources. Through these assessments, we aim to unravel the significance, credibility, and real-world ramifications of the identified limitations and constraints, equipping stakeholders with the insights necessary for responsible development, informed decision-making, and ethical utilization of ChatGPT and similar AI-driven conversational systems.

As we navigate the intricate intersection of artificial intelligence and human communication, this annotated bibliography strives to be a beacon of knowledge and discernment. It is our hope that the insights and perspectives contained within these annotated sources will not only illuminate the path toward harnessing ChatGPT's strengths but also guide us in addressing and mitigating its limitations and constraints effectively.

There are a large variety of types of limitations and constraints of ChatGPT that have the potential to advance the research on ChatGPT's capabilities, applications, and performances in both academic and professional environments. This annotated bibliography includes the following more interesting and important types that have caught attention of scholars, researchers and professionals.

### **Vulnerability to malicious use:**

Concerns about potential exploitation for malicious purposes take precedence due to the wide-ranging implications of misuse.

### **Biases and fairness issues:**

Addressing biases is crucial for promoting fairness and equity in AI applications, making it a high-priority consideration.

### **Challenge in upholding ethical standards:**

Ensuring adherence to ethical standards is paramount for responsible AI deployment, necessitating continuous monitoring and oversight.

### **Potential for misinformation propagation:**

Preventing the spread of misinformation is critical, given the impact on public understanding and the potential consequences of inaccurate information.



**Lack of inherent moral understanding:**

The absence of genuine moral understanding highlights the need for AI systems to better navigate nuanced ethical dilemmas.

**Inability to discern harmful Intent:**

The model's difficulty in discerning harmful intent raises ethical concerns, particularly in preventing content that could encourage harm or illegal activities.

**Need for user verification:**

Ensuring user identity verification becomes more important, especially to prevent impersonation and misuse by malicious actors.

**Insufficient safeguards against harmful content:**

Strengthening safeguards to consistently prevent the generation of harmful or offensive material is a priority to uphold user safety.

**No inherent sense of user well-Being:**

Considering user well-being becomes crucial, emphasizing the need for AI systems to prioritize positive user experiences.

**Security risks in model deployment:**

Implementing robust security measures during model deployment is essential to prevent unauthorized access or manipulation.

**Algorithmic bias and explainability:**

Addressing algorithmic biases and enhancing explainability are key to fostering trust and accountability in AI applications.

**Limited task-specific performance:**

The need for task-specific improvements underscores the importance of enhancing ChatGPT's performance in various applications.

**Challenge in adapting to new information:**

The model's limitation in adapting to new information necessitates ongoing updates to maintain accuracy and reliability.

**Usability and satisfaction concerns:**

User satisfaction and usability concerns, including issues related to reliability, are important for fostering positive interactions.

**Potential for overreliance on AI:**

Awareness of the potential for overreliance on AI emphasizes the importance of informed decision-making and responsible usage.

**2 Annotated bibliography**



Presented in the following section is the annotated bibliography for the academic studies on the limitations and constraints of ChatGPT, alphabetically arranged by the last name of the first author of the papers published before November 2023.

**Abdulai, AF & Hung, L. 2023. Will ChatGPT Undermine Ethical Values in Nursing Education, Research, and Practice. *Nurs. Inq.* e12556–e12556.**

In this commentary article, the authors, who are nursing researchers specializing in digital health technology and AI, express concerns that extend beyond the safety implications of ChatGPT. They are also wary of its ability to encompass the ethical values, principles, and fundamental doctrines that underpin the unique field of nursing. Recent editorials and commentaries in specific nursing journals have discussed the potential benefits, limitations, and risks associated with integrating ChatGPT into nursing education and practice. In this article, the authors further elaborate on these discussions by considering how the adoption of ChatGPT might potentially erode the very values, principles, and foundational assumptions that support nursing research, education, and practice. They aim to prompt discussions on the current and emerging trends in ChatGPT utilization within the nursing domain. Their goal is to actively involve nursing scholars, educators, and practitioners across international networks in conversations regarding nursing paradigms in the context of technological integration.

**Alawida, M; Mejri, S; Mehmood, A; Chikhaoui, B & Isaac Abiodun, O. 2023. A Comprehensive Study of ChatGPT: Advancements, Limitations, and Ethical Considerations in Natural Language Processing and Cybersecurity. *Information*, 14(8), 462.**

This research undertakes an exhaustive examination of ChatGPT, a cutting-edge language model reshaping generative text. We conduct a thorough analysis encompassing its architecture, training data, and evaluation metrics while tracing its evolution and improvements over time. The paper scrutinizes the capabilities and constraints of ChatGPT in various natural language processing (NLP) tasks, spanning language translation, text summarization, and dialogue generation. Additionally, a comparative evaluation against other language generation models is presented, shedding light on ChatGPT's versatility in diverse tasks. Ethical and privacy considerations related to ChatGPT are addressed, accompanied by insights into potential mitigation strategies. The research delves into the role of ChatGPT in cyberattacks, emphasizing associated security risks. Finally, the paper explores the wide-ranging applications of ChatGPT across industries, evaluating its performance across languages and domains. In essence, this paper offers a comprehensive exploration of ChatGPT's profound impact on the NLP field.

**AlZu'bi, S; Mughaid, A; Quiam, F & Hendawi, S. 2022. Exploring the capabilities and limitations of ChatGPT and alternative big language models. In: *Artificial Intelligence and Applications*.**

This paper delves into the phenomenon of ChatGPT, an AI-powered chatbot developed by OpenAI, which has garnered immense popularity since its public launch in November 2022. Renowned for its versatility in writing essays, emails, poems, and even computer code, ChatGPT serves as a valuable tool for professionals across diverse fields. However, it is crucial to note that ChatGPT's responses are generated by a GAN, occasionally deviating from reality. The primary objective of this paper is to construct a text classification model for a chatbot using Python. The model undergoes training on a dataset comprising customer responses to a survey and their corresponding class labels. Various classifiers, including Naive Bayes, Random Forest, Extra Trees, and Decision Trees, are trained and tested. The results reveal that the Extra Trees classifier outperforms others with an accuracy of 90%. The study emphasizes the significance of text preprocessing and the selection of appropriate classifiers for text classification tasks in building an efficient chatbot. Furthermore, the paper explores the capabilities and limitations of ChatGPT and its alternatives in the year 2023. A comprehensive overview of the alternatives' performance is presented, culminating in a discussion on the future directions of large language models and their broader impact on society and technology.

**Alzboon, MS; Qawasmeh, S; Alqaraleh, M; Abuashour, A; Bader, AF & Al-Batah, M. 2023. Pushing the envelope: investigating the potential and limitations of ChatGPT and artificial intelligence in advancing**

**computer science research. In: 2023 3rd International Conference on Emerging Smart Technologies and Applications (eSmarTA), pp 1-6. IEEE.**

This paper explores the transformative impact of artificial intelligence (AI) on computer science, offering new avenues for data analysis, modeling, and prediction. Among the AI models, ChatGPT, based on deep neural networks, exhibits promise in natural language processing, image recognition, and data analysis. Nevertheless, the integration of ChatGPT and other AI models into scientific research encounters challenges and limitations. The primary focus of this study is to analyze and address the constraints of ChatGPT and AI in computer science research. Specifically, we target two key objectives. Firstly, we address Domain expertise restrictions, recognizing the potential lack of specialized subject knowledge in ChatGPT and similar language models. This deficiency poses challenges for researchers in specific domains. Our investigation delves into methodologies to enhance the integration of domain knowledge into AI models, facilitating more precise and contextually relevant predictions. Secondly, we tackle interpretability difficulties inherent in AI models, a significant hurdle impeding researchers' comprehension of algorithmic decision-making processes. The study proposes strategies to enhance interpretability, providing researchers with insights into the mechanisms behind AI model predictions. The overarching goal is to foster the ethical and responsible use of AI in scientific research by addressing these objectives. The paper examines diverse strategies aimed at refining the representation of domain knowledge within AI models and enhancing their interpretability. Through these efforts, we aim to empower researchers to harness the benefits of ChatGPT and AI while mitigating their inherent limitations.

**Amaro, I; Della Greca, A; Francese, R; Tortora, G & Tucci, C. 2023. AI unreliable answers: A case study on ChatGPT. In: International Conference on Human-Computer Interaction, pp 23-40. Cham: Springer Nature Switzerland.**

In this study, the researchers examined the reliability of answers from ChatGPT and how expert users, such as Computer Science students, perceive its reliability. Initially, they assessed the accuracy of ChatGPT's responses by testing its narrative, problem-solving, searching, and logic capabilities, providing examples of the answers obtained. Subsequently, a user study involving 15 participants familiar with the chatbot was conducted. The participants posed a predetermined set of queries, generating both correct and incorrect responses, and their satisfaction levels were collected. Results indicated that, despite occasional unreliability in the current version of ChatGPT, users still express an intention to utilize it. Consequently, it is recommended to use the current version of ChatGPT in conjunction with human verification and interpretation for enhanced reliability.

**Ayub, I; Hamann, D; Hamann, CR & Davis, MJ. 2023. Exploring the potential and limitations of chat generative pre-trained transformer (ChatGPT) in generating board-style dermatology questions: a qualitative analysis. Cureus, 15(8).**

This paper explores the constraints of Chat Generative Pre-trained Transformer (ChatGPT), an OpenAI language model, when used as a study tool in dermatology. Employing ChatPDF, an application integrating PDF files with ChatGPT, the study generated American Board of Dermatology Applied Exam (ABD-AE)-style questions from Journal of the American Board of Dermatology articles. A qualitative analysis by two board-certified dermatologists assessed 40 questions for accuracy, complexity, and clarity, revealing that only 16 (40%) were deemed suitable for ABD-AE study preparation. Limitations observed included low complexity, lack of clarity, and inaccuracies, highlighting ChatGPT's challenges in grasping domain-specific knowledge in dermatology. The model's inability to understand context and generate high-quality distractors, coupled with the absence of image generation capabilities, further diminishes its utility. The study underscores that while ChatGPT may assist in generating simple questions, it cannot replace the expertise of dermatologists and medical educators in crafting high-quality, board-style questions that effectively assess candidates' knowledge and reasoning abilities.

**Azaria, A. 2022. ChatGPT usage and limitations. hal-03913837.**

This research delves into the multifaceted utility of large language models across diverse domains such as conversational agents, education, and explainable AI. One notable instance is ChatGPT, a substantial language



model crafted by OpenAI specifically designed as a conversational agent. Given its extensive training on vast datasets, ChatGPT's responses are inevitably influenced by the characteristics inherent in the training data. In this paper, a distinctive bias in ChatGPT related to the usage of digits in numbers is emphasized. Notably, a remarkably high correlation is unveiled between the frequency of digits generated by ChatGPT and the preferred numbers of humans, with the most frequently generated digit aligning with the widely favored human number, 7. Additionally, we explore the advantageous aspects of ChatGPT as a conversational agent and deliberate on certain limitations inherent in its design.

**Bang, J & Park, G. 202).** Uprising of ChatGPT and ethical problems. *Robotics & AI Ethics*, pp 1-11.

This paper investigates ethical considerations arising from the introduction of ChatGPT, focusing on moral competence and proposing ethical solutions. Employing a comprehensive review of literature and media reports from both Eastern and Western perspectives, the study analyzes ChatGPT's technology, operational methods, and associated ethical concerns. The examination reveals potential issues such as plagiarism, copyright infringement, test fairness, criminal misuse, social stereotypes, privacy invasion, security vulnerabilities, diminished critical thinking, and erosion of genuine human relationships. Evaluation based on moral competence highlights the feasibility of implementing moral identity, moral sensitivity, and moral practice, but underscores limitations in moral judgment. To address these ethical challenges, the study advocates for a utilitarian approach, emphasizing decisions that maximize societal benefits while minimizing harm. The conclusion emphasizes the necessity of embedding ethical perspectives into responsible AI toolkits and frameworks for balanced decision-making in the design, development, adoption, deployment, maintenance, and evolution of ChatGPT. Furthermore, it underscores the importance of a multi-stakeholder approach for fostering a socially beneficial AI society.

**Barrot, JS. 2023.** Using ChatGPT for second language writing: pitfalls and potentials. *Assessing Writing*, 57, 100745.

This research delves into the contemporary developments in artificial intelligence, specifically focusing on the utilization of chatbots as valuable aids in language learning. Among these tools, ChatGPT stands out, offering users a natural and human-like interactive learning experience. Despite its potential as an effective language tutor and source of linguistic input, concerns have been raised by academics regarding its influence on writing pedagogy and academic integrity. This tech review seeks to comprehensively examine both the potential advantages and challenges associated with integrating ChatGPT into second language (L2) writing contexts. The paper concludes by providing recommendations for enhancing L2 writing classroom practices in light of these findings.

**Belgas, I; Hilal, I & Afifi, N. 2023.** Large language models in e-learning: advantages and limitations of ChatGPT. In: *COC2023*.

This paper explores the ever-evolving landscape of education in the contemporary era, propelled by the advent of new artificial intelligence technologies. Intelligent systems, including Large Language Models (LLMs) like ChatGPT, play a crucial role in enhancing the learning process by leveraging knowledge from tutors and transferring it to learners. Distinguished by its capability to effectively interact with human queries through Natural Language Processing (NLP) tools, ChatGPT comprehends human intent and generates responses.

The primary objective of this work is to identify, collect, and synthesize numerous research endeavors that delve into the utilization of LLMs in the field of education. Subsequently, a comparative study is conducted to illuminate the advantages and disadvantages gleaned from these works, providing insights that can fuel new research ideas for future endeavors in this domain.

**Bengio, Y. 2023.** AI and catastrophic risk. *Journal of Democracy*, 34(4): 111-121.

This paper explores the growing concerns about the potential threats to democracy stemming from the release of large language models such as Chat-GPT and GPT-4 by OpenAI. The focus is on the concept of "rogue AIs," which refers to powerful and hazardous artificial intelligences capable of pursuing harmful objectives independently of human intentions. The essay advocates for proactive measures to address the risks posed by these rogue AIs to

democracy and global geopolitical stability. The author emphasizes the need for international collaboration and suggests that research into safe and defensive AI technologies should be conducted through a multilateral network involving research laboratories from various countries.

**Borji, A. 2023. A categorical archive of ChatGPT failures. *arXiv preprint arXiv:2302.03494*.**

In this paper, the author explored various drawbacks of ChatGPT and emphasized its limitations. While it exhibits impressive capabilities in specific tasks, there is a necessity for further improvements to excel in areas such as logical reasoning, solving mathematical problems, mitigating biases, and more. Presently, ChatGPT remains susceptible to these issues. It is uncertain whether these limitations can be resolved given the ambiguous capabilities of current technology. The reliability and credibility of both ChatGPT and upcoming models are also under scrutiny. While this research thoroughly evaluates ChatGPT, it does possess certain shortcomings that should be addressed in subsequent studies, as outlined below:

- The extent to which ChatGPT memorizes versus understands what it generates is still unknown.
- The degree to which ChatGPT possesses commonsense and ways to enhance it is uncertain.
- ChatGPT seems limited in its ability to generate creative solutions to novel problems, particularly in unsolved mathematics.
- ChatGPT lacks a mechanism to indicate when it is uncertain about its answers.
- ChatGPT's responses are inconsistent and can sometimes be contradictory.
- The existence of biases in the training data is one of the primary ethical concerns surrounding the use of ChatGPT.
- ChatGPT's capacity to comprehend and address intricate inquiries is restricted.
- The use of ChatGPT has the potential to impersonate a trusted person.
- Unintended consequences are often challenging to predict.

**Carvalho, I & Ivanov, S. 2023. ChatGPT for tourism: applications, benefits and risks. *Tourism Review*.**

This paper explores the applications, advantages, and potential risks of ChatGPT and large language models within the tourism sector, aiming to establish a research agenda for further investigation. By reviewing existing literature on ChatGPT, large language models, and artificial intelligence, the paper identifies areas where ChatGPT could be applied for various stakeholders in the tourism industry. The findings suggest that ChatGPT and similar models are poised to significantly impact tourism processes, improving customer service, and enhancing efficiency in operations. While potential negative consequences on human resources are acknowledged, the overall impact is anticipated to be positive, particularly in empowering and assisting tourism employees. This study contributes to the field by providing insights into the potential implications of ChatGPT in tourism and hospitality, representing one of the initial explorations in this area.

**Cheng, HW. 2023. Challenges and limitations of ChatGPT and artificial intelligence for scientific research: a perspective from organic materials. *AI*, 4(2): 401-405.**

This paper has concentrated on delineating the challenges and limitations encountered by ChatGPT within the realm of organic materials research, with the goal of offering insights into the obstacles faced by researchers in this particular field. The author has discerned that ChatGPT exhibits a deficiency in furnishing valuable information, attributable to three potential limitations in its responses: (i) computational calculation constraints, (ii) the propensity for inaccurate or false facts, and (iii) shortcomings in inferential capability. These limitations have the potential to engender misunderstandings or misinterpretations. The author has also scrutinized various ethical considerations with potential concerns, including the risk of biases in the generated content stemming from biases present in user-provided data, apprehensions regarding privacy and security when handling sensitive or unpublished data, and the possibility that the generated content may be subject to intellectual property laws, thereby giving rise to legal complexities and the risk of copyright violations. Additionally, accountability assumes significance as a critical aspect to be taken into account in the deployment of AI tools such as ChatGPT in scientific research.



**Chowdhury, MNUR & Haque, A. 2023. ChatGPT: its applications and limitations. In: 2023 3rd International Conference on Intelligent Technologies (CONIT), pp 1-7. IEEE.**

This paper examines ChatGPT, a cutting-edge language model that has garnered widespread attention for its remarkable ability to generate human-like responses to natural language inputs. Trained on an extensive corpus of text data, ChatGPT delivers swift and accurate responses across a diverse array of user queries. However, alongside its impressive capabilities, the model grapples with certain limitations. A primary challenge lies in ChatGPT's propensity to generate biased or inappropriate responses based on its training data. Furthermore, its reliance on statistical patterns in text data may pose difficulties in handling more nuanced or complex language tasks. Despite these limitations, ChatGPT marks a significant advancement in conversational AI, with vast potential applications ranging from chatbots and customer service to content creation and education. As technology progresses, it is anticipated that more sophisticated language models will emerge, surpassing the capabilities of ChatGPT and addressing some of its limitations. This evolution promises even greater strides in the development of conversational AI, opening up new possibilities for enhanced applications in various fields.

**Cotton, DR; Cotton, PA & Shipway, JR. 2023. Chatting and cheating: ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 1-12.**

This paper investigates the application of artificial intelligence, particularly focusing on ChatGPT, in the realm of academia—a topic of considerable interest in the education sector. While ChatGPT brings about advantages such as heightened student engagement, improved collaboration, and enhanced accessibility, it also introduces apprehensions regarding academic integrity and plagiarism. The study thoroughly explores the opportunities and challenges associated with integrating ChatGPT into higher education, weighing the potential risks and rewards of these AI tools. Additionally, it scrutinizes the complexities of identifying and preventing academic dishonesty and proposes strategies that universities can adopt to ensure ethical and responsible usage of such tools. These strategies encompass the formulation of policies and procedures, provision of training and support, and the implementation of diverse methods for detecting and thwarting cheating. In conclusion, the paper asserts that despite the dual nature of opportunities and challenges presented by AI in higher education, universities can effectively navigate these concerns through a proactive and ethical approach to the incorporation of these tools.

**Currie, GM. 2023. Academic integrity and artificial intelligence: is ChatGPT hype, hero or heresy? In: *Seminars in Nuclear Medicine*. WB Saunders.**

This paper examines the impact of artificial intelligence advancements on academic integrity in higher education and scientific writing, with a particular focus on the role of ChatGPT in addressing associated challenges. Unlike traditional algorithms, ChatGPT, powered by GPT-3.5, stands out for its ability to generate precise and realistic responses to questions in real-time. While holding promise in various applications, this chatbot faces notable constraints when applied to nuclear medicine and radiology. In the context of nuclear medicine, ChatGPT's susceptibility to errors and the potential generation of false information pose risks to professionalism, ethics, and integrity. These limitations compromise the reliability of ChatGPT, falling short of the expected standards required in these fields. However, despite these challenges, there are promising applications for ChatGPT in nuclear medicine within educational, clinical, and research domains. Integrating ChatGPT into practice requires a reevaluation of norms and a reshaping of expectations regarding information reliability. This paper explores both the constraints and potential applications of ChatGPT in nuclear medicine, emphasizing the need for careful consideration and adaptation of existing practices to fully harness its capabilities.

**Dave, T; Athaluri, SA & Singh, S. 2023. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. *Frontiers in Artificial Intelligence*, 6, 1169595.**

This paper conducts an in-depth analysis of the advantages, limitations, ethical considerations, future prospects, and practical applications associated with ChatGPT and artificial intelligence (AI) in the healthcare and medical sectors. Operating as an advanced language model, ChatGPT utilizes deep learning techniques to generate human-like responses to natural language inputs. As part of the generative pre-training transformer (GPT) model family



developed by OpenAI, ChatGPT stands as one of the largest publicly available language models. Capable of capturing the nuances of human language, ChatGPT demonstrates its potential applications in the medical domain. From identifying research topics to aiding professionals in clinical and laboratory diagnosis, its versatility extends to keeping healthcare practitioners informed about updates and developments in their fields. The creation of virtual assistants to support patients in managing their health is another crucial application of ChatGPT in medicine. Despite the promising applications, the integration of ChatGPT and other AI tools in medical writing raises ethical and legal concerns. These include potential copyright infringement, medico-legal complications, and the imperative need for transparency in AI-generated content. In conclusion, while ChatGPT exhibits numerous potential applications in the medical and healthcare sectors, this paper thoroughly examines the associated limitations and ethical considerations. The future prospects in medicine and healthcare are also explored, emphasizing the need for a balanced and conscientious approach to the integration of AI technologies.

**Day, T. 2023. A preliminary investigation of fake peer-reviewed citations and references generated by ChatGPT. *The Professional Geographer*, 1-4.**

The author analyzes the academic citations and references generated by the ChatGPT artificial intelligence (AI) chatbot and reveals the citations and references are in fact, fake. They are clearly generated by a predictive process rather than known facts. This suggests that early optimism regarding this technology for assisting in research could be misplaced, and that student misuse of the chatbot can be detected by the identification of fake citations and references. Despite these problems, the technology could have application in the writing of course materials for lower level undergraduate courses that do not necessarily require references. Subject matter expertise is required, however, to identify and remove incorrect information. The need to identify incorrect information provided by an AI chatbot is a skill that students will also increasingly need.

**Derner, E & Batistič, K. 2023. Beyond the safeguards: exploring the security risks of ChatGPT. *arXiv preprint arXiv:2305.08005*.**

This paper addresses the escalating concerns surrounding the safety, security risks, and ethical considerations associated with the widespread use of large language models (LLMs), exemplified by ChatGPT. The focus is on exploring various security risks linked to ChatGPT, encompassing malicious text and code generation, disclosure of private data, provision of fraudulent services, information gathering, and the generation of unethical content. Through an empirical study, we evaluate the effectiveness of ChatGPT's content filters and scrutinize potential methods to circumvent these safeguards, highlighting persistent ethical implications and security risks within LLMs even when protective measures are implemented. Drawing from a qualitative analysis of security implications, we propose potential strategies to mitigate these risks, providing valuable insights for researchers, policymakers, and industry professionals dealing with the intricate security challenges presented by LLMs like ChatGPT. This study contributes to the ongoing discourse surrounding the ethical and security considerations associated with LLMs, emphasizing the necessity for continuous research in this domain.

**Deshpande, A; Murahari, V; Rajpurohit, T; Kalyan, A & Narasimhan, K. 2023. Toxicity in ChatGPT: Analyzing persona-assigned language models. *arXiv preprint arXiv:2304.05335*.**

Large language models (LLMs) have demonstrated remarkable capabilities, extending beyond the natural language processing (NLP) community and finding application in various sectors such as healthcare, therapy, education, and customer service. As these models are utilized by users with critical information needs, such as students or patients interacting with chatbots, ensuring the safety of these systems becomes paramount. Therefore, it is essential to have a clear understanding of the capabilities and limitations of LLMs. To address this, the authors conducted a systematic evaluation of toxicity in over half a million generations of ChatGPT, a popular dialogue-based LLM. Our findings indicate that adjusting the system parameter of ChatGPT by assigning it a persona, such as that of the boxer Muhammad Ali, significantly increases the toxicity of its responses. Depending on the assigned persona, the toxicity of ChatGPT can rise by up to 6x, leading to outputs that involve incorrect stereotypes, harmful dialogue, and hurtful opinions. This has the potential to be defamatory to the persona and harmful to unsuspecting users. Additionally, the authors observed troubling patterns where specific entities, such as certain races, are targeted more than others (3x



more), irrespective of the assigned persona, revealing inherent discriminatory biases in the model. The authors hope that these findings prompt the broader AI community to reconsider the effectiveness of current safety measures and encourage the development of improved techniques that ensure the robustness, safety, and trustworthiness of AI systems.

**Elyoseph, Z & Levkovich, I. 2023. Beyond human expertise: the promise and limitations of ChatGPT in suicide risk assessment. *Frontiers in psychiatry*, 14.**

This paper explores the potential of ChatGPT, an artificial intelligence language model developed by OpenAI, in the realm of mental health, with a specific focus on suicide prevention. While ChatGPT holds theoretical promise, its clinical abilities in addressing this significant mental health concern remain untested. To bridge this knowledge gap, the study aims to compare ChatGPT's assessments of mental health indicators with those of mental health professionals, particularly in the context of a hypothetical case study centered on suicide risk assessment. In this investigation, ChatGPT was tasked with evaluating a text vignette describing a hypothetical patient exhibiting varying levels of perceived burdensomeness and thwarted belongingness. The assessments provided by ChatGPT were then compared to the norms established by mental health professionals. The results revealed that, across all conditions, ChatGPT rated the risk of suicide attempts lower than mental health professionals did. Additionally, in most conditions, ChatGPT rated mental resilience lower than the established norms. These findings suggest that individuals, including gatekeepers, patients, or mental health professionals relying on ChatGPT for evaluating suicide risk or enhancing decision-making may receive inaccurate assessments that underestimate the actual risk of suicide.

**Esmailzadeh, Y. 2023. Potential risks of ChatGPT: implications for counterterrorism and international security. *International Journal of Multicultural and Multireligious Understanding (IJMMU)* Vol 10.**

This paper delves into the examination of potential risks associated with ChatGPT, a language generation model developed by OpenAI, concerning counterterrorism and international security. Despite its widespread adoption, ChatGPT's strengths and weaknesses raise practical and ethical concerns that warrant careful consideration. The escalating use of modern communication technologies by terrorists has intensified the complexity of intelligence challenges, making critical data access increasingly difficult for intelligence agencies. Terrorist groups leverage new technologies and implement enhanced operational security measures, posing challenges to sophisticated intelligence collection operations. Moreover, the advent of artificial intelligence amplifies the threat of cyberattacks and espionage, as individuals and groups exploit these technologies. The study analyzes ChatGPT's potential risks across four critical domains: the implications of artificial intelligence on future threats and international security, the influence of ChatGPT on cyberterrorism and artificial intelligence, the hazards associated with fragmented and secondary information in violence and sabotage operations, and the utilization of psychological warfare against targets. By offering insights and recommendations, this research aims to guide policymakers and researchers in effectively employing artificial intelligence technology while minimizing potential risks.

**Ferrara, E. 2023. Should ChatGPT be biased? challenges and risks of bias in large language models. *arXiv preprint arXiv:2304.03738*.**

This paper examines the challenges and risks associated with biases present in large-scale language models like ChatGPT. It delves into the sources of biases, including aspects such as training data, model specifications, algorithmic constraints, product design, and policy decisions. The paper also addresses ethical concerns stemming from the unintended consequences of biased model outputs. Additionally, it assesses potential opportunities to mitigate biases, the inevitability of certain biases, and the implications of deploying these models in various applications like virtual assistants, content generation, and chatbots. Furthermore, the paper analyzes existing approaches to identify, quantify, and address biases in language models. It underscores the necessity for a collaborative, multidisciplinary effort to develop AI systems that are more equitable, transparent, and responsible. The overarching goal of this paper is to stimulate thoughtful dialogue within the artificial intelligence community, prompting researchers and developers to contemplate the role of biases in generative language models and the ongoing pursuit of ethical AI.

**Frieder, S; Pinchetti, L; Griffiths, RR; Salvatori, T; Lukasiewicz, T; Petersen, PC & Berner, J. 2023. Mathematical capabilities of ChatGPT. *arXiv preprint arXiv:2301.13867*.**

This research delves into the mathematical capabilities of two versions of ChatGPT (released on January 9, 2023, and January 30, 2023) as well as GPT-4. To assess their performance, The authors employed a novel methodology, testing them on publicly available datasets and custom-created datasets. Unlike formal mathematics, which benefits from extensive databases of formal proofs (e.g., the Lean Mathematical Library), existing datasets for natural-language mathematics, used for evaluating language models, are limited, often covering only elementary mathematics. In response, The authors have released two new datasets, GHOSTS and miniGHOSTS, curated by active mathematics researchers. These datasets aim to (1) encompass graduate-level mathematics, (2) offer a comprehensive overview of language models' mathematical capabilities, and (3) assess various dimensions of mathematical reasoning. The study also explores whether ChatGPT and GPT-4 can serve as helpful assistants to professional mathematicians, emulating scenarios encountered in their daily activities. The authors conducted a thorough evaluation using fine-grained performance metrics, representing the most detailed assessment for advanced mathematics to date. The findings indicate that ChatGPT excels as a mathematical assistant for factual inquiries, effectively functioning as a mathematical search engine and knowledge base interface. GPT-4, while suitable for undergraduate-level mathematics, falls short when confronted with the challenges of graduate-level difficulty. Despite media reports emphasizing GPT-4 and ChatGPT's success in solving exams (potentially influenced by selection bias), their overall mathematical performance is considerably below that of a graduate student. Consequently, if the aim is to rely on ChatGPT for success in a graduate-level math exam, one would be better off consulting an average peer!

**Gao, J; Zhao, H; Yu, C & Xu, R. 2023. Exploring the feasibility of ChatGPT for event extraction. *arXiv preprint arXiv:2303.03836*.**

This paper addresses the fundamental task of event extraction in natural language processing, involving the identification and extraction of information about events mentioned in text. Despite being a challenging task due to the scarcity of annotated data, the emergence of large language models (LLMs) like ChatGPT provides an opportunity to tackle language tasks without the need for specific datasets and fine-tuning. While ChatGPT has demonstrated impressive results in tasks like machine translation, text summarization, and question answering, it encounters challenges in complex tasks such as event extraction. Unlike other tasks, event extraction requires the model to be provided with a complex set of instructions defining all event types and their schemas. To assess the feasibility of ChatGPT for event extraction and the challenges it poses, a series of experiments were conducted. The results reveal that, on average, ChatGPT performs at only 51.04% of the level of a task-specific model like EEQA in long-tail and complex scenarios. Usability testing experiments indicate that ChatGPT lacks robustness, and continuous refinement of the prompt does not lead to stable performance improvements, potentially resulting in a poor user experience. Additionally, ChatGPT exhibits high sensitivity to different prompt styles.

**Gladstone, R. 2023. Using ChatGPT in the classroom: opportunities, limitations, and ethical considerations. *Spicer Adventist University Research Articles Journal*, 2(1): 17-25.**

This paper marks a significant juncture in the landscape of open artificial intelligence (AI), specifically in November 2022. The educational community witnessed a notable shift as a sophisticated chatbot, ChatGPT, developed by OpenAI, rapidly amassed a million users within a month. Delving into the historical trajectory of ChatGPT, this paper aims to unravel its capabilities and scrutinize the potential benefits and limitations of integrating it as a teaching tool in the classroom. Moreover, the paper undertakes a thorough examination of the ethical concerns and privacy implications associated with the utilization of models like ChatGPT. The overarching goal is to provide educators with valuable insights, equipping them with the knowledge needed to effectively incorporate ChatGPT into their teaching practices while maintaining a vigilant awareness of ethical considerations.

**Gravel, J; D'Amours-Gravel, M & Osmanliu, E. 2023. Learning to fake it: limited responses and fabricated references provided by ChatGPT for medical questions. *Mayo Clinic Proceedings: Digital Health*, 1(3): 226-234.**

This study aimed to assess the quality of answers and references provided by ChatGPT for medical questions. Three researchers posed 20 medical questions to ChatGPT, requesting corresponding references. Medical experts, who were the corresponding authors of the 20 articles from which the questions were derived, evaluated the responses using a verbal numeric scale ranging from 0% to 100%. Initially planning to assess three references per response for pertinence, this was modified due to preliminary findings indicating the fabrication of most references by ChatGPT. The study, conducted in February 2023, revealed responses ranging from 53 to 244 words, with 2 to 7 references per answer. Feedback from 17 of the 20 invited raters indicated limited response quality, with a median score of 60%. Additionally, the raters identified major (n=5) and minor (n=7) factual errors in 17 responses. Of the 59 references evaluated, 41 (69%) were found to be fabricated, despite appearing real. Fabricated citations often included names of authors with relevant previous publications, seemingly pertinent titles, and a credible journal format. This underscores the need for users to scrutinize ChatGPT's provided references, particularly when integrating them into medical manuscripts.

**Hadi, MU; Qureshi, R; Shah, A; Irfan, M; Zafar, A; Shaikh, MB & Mirjalili, S. 2023. A survey on large language models: applications, challenges, limitations, and practical usage. *TechRxiv*.**

This paper delves into the expansive realm of computerized language processing, focusing on the transformative presence of Large Language Models (LLMs). Within the domain of artificial intelligence, LLMs wield remarkable capabilities, demonstrating an unparalleled capacity to grasp intricate linguistic nuances and generate coherent and contextually fitting responses. Exploring their historical roots, architecture, training methodologies, applications, and challenges, this survey paper aims to provide a comprehensive understanding of LLMs. Commencing with an exploration of fundamental concepts in generative AI and an analysis of the architecture of generative pretrained transformers (GPT), the paper proceeds to trace the evolution of LLMs over time. It delves into the various training methods employed in shaping these models. Moving beyond theoretical foundations, the survey outlines the diverse applications of LLMs across fields such as medicine, education, finance, and engineering, elucidating how they contribute to solving real-world problems. As LLMs increasingly shape the landscape of AI, this paper scrutinizes the associated challenges in deploying them in practical scenarios. Addressing ethical considerations, model biases, interpretability, and computational resource requirements, the survey sheds light on the complexities involved. It further underscores techniques aimed at bolstering the robustness and controllability of LLMs, while also tackling issues related to bias, fairness, and generation quality.

In conclusion, the paper anticipates the future trajectory of LLM research, emphasizing the need to overcome challenges to enhance the reliability and utility of these models. Designed as a valuable resource for researchers, practitioners, and enthusiasts, this survey consolidates state-of-the-art knowledge in the field, serving as a compass for advancements in the development and application of LLMs across a spectrum of real-world contexts.

**Haverkamp, W; Tennenbaum, J & Strodthoff, N. 2023. ChatGPT fails the test of evidence-based medicine. *European Heart Journal-Digital Health*, ztad043.**

This research underscores the crucial need to acknowledge the limited documentation of the training data and methods employed by ChatGPT, much of which has been shrouded in secrecy. Notably, ChatGPT, when directly questioned, lacks the ability to specify the incorporation of medical guidelines into its training. The standard of evidence-based medicine necessitates transparency in sourcing and a meticulous selection of those sources, highlighting the importance of clarity regarding the utilized guidelines for comprehensibility. Furthermore, the imperative for information currency is emphasized, with the knowledge and training data for ChatGPT having a last cut-off date in September 2021. In a comprehensive assessment, ChatGPT reveals numerous gaps and limitations that warrant an in-depth discussion among medical experts. This discourse should pave the way for an objective evaluation of ChatGPT's clinical relevance and applicability. Unfortunately, discussions surrounding the capabilities of chatbots often prove ambiguous and problematic. It is unlikely, even with improvements, that ChatGPT will achieve the competence, reliability, and trustworthiness demanded by evidence-based medicine. Perhaps, future



advancements may introduce more specialized systems that address the concerns raised in this study. Only time will unveil the trajectory of these developments.

**Hou, Y; Yeung, J; Xu, H; Su, C; Wang, F & Zhang, R. 2023. From answers to insights: unveiling the strengths and limitations of ChatGPT and biomedical knowledge graphs. *medRxiv*, 2023-06.**

This paper examines the performance of Large Language Models (LLMs) in the context of natural language processing tasks, with a particular focus on their application in the biomedical domain. Despite their exceptional capabilities in language generation and knowledge acquisition from unstructured text, LLMs encounter limitations in this specialized field, leading to inaccuracies and inconsistencies. In response to these challenges, Biomedical Knowledge Graphs (BKGs) have emerged as valuable tools for structured information representation and organization.

The study evaluates the capabilities of ChatGPT, particularly the GPT-4.0 version, in comparison to both GPT-3.5 and existing BKGs concerning question answering, knowledge discovery, and reasoning. The results indicate that while ChatGPT excels in providing existing information, BKGs demonstrate superior reliability in information delivery. However, ChatGPT exhibits limitations in novel discoveries and reasoning, especially in establishing structured links between entities when compared to BKGs. To address these shortcomings, the paper suggests that future research should focus on integrating LLMs and BKGs, capitalizing on their complementary strengths. This integrated approach holds the potential to optimize task performance, mitigate risks, and advance knowledge in the biomedical field, contributing to overall well-being.

**Huallpa, JJ. 2023. Exploring the ethical considerations of using Chat GPT in university education. *Periodicals of Engineering and Natural Sciences*, 11(4): 105-115.**

This research delves into the ethical quandaries associated with the integration of Chat GPT into higher education, particularly focusing on the context of Latin American institutions of higher learning. A survey involving 220 participants through an online questionnaire aimed to understand their experiences and motivations in utilizing AI-powered conversational agents. Descriptive statistics provided a demographic overview of the participants. This study not only establishes a foundation for further exploration but also uncovers hidden meanings within observed phenomena, proposing potential solutions to the identified issues. It examines how AI systems and chatbots can complement human knowledge and judgment, alongside potential drawbacks. Results indicated that participants perceived Chat GPT integration as moderately accessible with moderately positive social attitudes. They acknowledged the value and responsibility of Chat GPT in creating personalized educational opportunities. Emphasizing the need for explicit institutional standards on privacy and data security, participants' reliance on AI was influenced by factors such as gender, age, accessibility, social attitude, opinions, personal experience, privacy and data security, institutional guidelines, and individualized learning, as revealed by regression analysis. The findings shed light on the complexities of integrating Chat GPT into Latin American higher education, influenced by individual beliefs, cultural norms, and ethical considerations. The adaptability of AI addresses the busy schedules of students, making resources readily available for success. Natural language processing models offer instant assistance through text chat, voice, or video. The research suggests that to fully grasp the ethical implications and guide the development of responsible implementation strategies, additional qualitative investigations, longitudinal studies, and comparative research across diverse contexts are necessary. Bridging these knowledge gaps will propel the field of conversational AI ethically and beneficially within educational settings.

**Hung, J & Chen, J. 2023. The benefits, risks and regulation of using ChatGPT in Chinese academia: a content analysis. *Social Sciences*, 12(7): 380.**

This paper investigates the major benefits and risks associated with Chinese students utilizing ChatGPT for academic activities and explores potential regulations for maintaining academic integrity and ethical standards in Chinese academic settings. Utilizing content analysis of 40 relevant newspaper articles retrieved through a Google news search with specific keywords, the study reveals a polarized opinion in China regarding ChatGPT's role in academic responsibilities. While some express concerns about academic cheating facilitated by ChatGPT, others advocate for integrating AI-powered technologies into academic learning to enhance the quality of academic outputs.



Plagiarism violations emerge as a primary concern among Chinese educators. The study suggests the need for increased regulations in Chinese academia, emphasizing the importance of teachers guiding students in fact-checking AI-generated content, incorporating proper citations and references, and approaching AI-powered tools with systematic and creative questioning to optimize intellectual outputs.

**Irvine, DJ; Halloran, LJ & Brunner, P. 2023. Opportunities and limitations of the ChatGPT advanced data analysis plugin for hydrological analyses. *Hydrological Processes*, 37(10): e15015.**

This paper delves into the transformative impact of artificial intelligence (AI) tools, exemplified by ChatGPT, on various industries, with a specific focus on the evolving landscape of data analyses. While recent discussions have emphasized the ethical considerations associated with AI-generated text, the ChatGPT Code Interpreter, now known as the Advanced Data Analysis plugin, showcases AI's expanding capabilities in generating, translating, and adapting computer scripts. Through four illustrative examples, we highlight ChatGPT's proficiency in prompt-based generation of Python scripts for hydrological dataset analysis. The examples encompass tasks such as converting R functions to Python for baseflow separation, translating MATLAB functions to Python for temperature data 'web scraping,' generating a correlation matrix for major ion groundwater data, and presenting river flow data boxplots. The paper explores the considerable opportunities, existing limitations, and concerns linked to the utilization of ChatGPT in the realm of hydrological data analyses.

**Kalla, D & Smith, N. 2023. Study and analysis of Chat GPT and its impact on different fields of study. *International Journal of Innovative Science and Research Technology*, 8(3).**

This paper delves into the revolutionary technology of ChatGPT, utilizing advanced artificial intelligence techniques to generate natural language responses in response to prompts or inputs. With applications spanning natural language processing, customer service, and content creation, this study conducts a comprehensive analysis of ChatGPT. It aims to unravel the origins of ChatGPT, elucidate its functioning, and assess its impact across various fields of study.

The investigation systematically examines the advantages and disadvantages of ChatGPT, probing into its inherent limitations and distinctive features. Moreover, the paper explores the ramifications of ChatGPT on diverse domains such as academics, cyber security, customer support, software development, jobs, and information technology. Additionally, it scrutinizes the potential applications of ChatGPT for researchers and scholars, contemplating its evolving role in shaping the landscape of artificial intelligence.

**Kenney, NM. 2023. A brief analysis of the architecture, limitations, and impacts of ChatGPT.**

This paper commences with an exploration of the technical architecture of ChatGPT and distinguishes its features from other large-scale artificial intelligence language models. Following this, we outline the present use cases of ChatGPT, accompanied by an examination of several existing limitations. Specifically, we scrutinize ChatGPT's challenges in exhibiting creativity, address issues related to the perpetuation of biases, and consider identification concerns. Conclusively, we assess some of the anticipated key impacts of ChatGPT, encompassing considerations of copyright and economic implications. It is noteworthy that this analysis is approached from a non-technical standpoint, aiming to illustrate how AI is progressively intertwining with everyday life.

**Khalil, M & Er, E. 2023. Will ChatGPT get you caught? rethinking of plagiarism detection. *arXiv preprint arXiv:2302.04335*.**

This study addresses the escalating concerns surrounding the integration of Artificial Intelligence (AI) technology, particularly chatbots, into education and its potential repercussions. The heightened accessibility and enhanced capabilities of advanced AI systems, exemplified by ChatGPT, have raised apprehensions, particularly regarding the use of these chatbots for generating academic essays in educational settings. The primary objective of this research is to investigate the originality of content produced by ChatGPT. Utilizing two widely recognized plagiarism detection tools, the study assesses the originality of 50 essays generated by ChatGPT across various topics. The results reveal that ChatGPT possesses a significant capacity to generate sophisticated and highly original text



outputs, often evading detection by conventional plagiarism check software. This capability raises concerns about students utilizing chatbots as an expedient route to academic success with minimal effort. Furthermore, ChatGPT's ability to verify its own generated essays, showcasing superior performance compared to traditional plagiarism-detection tools, adds an additional layer to the plagiarism-checking process. The paper emphasizes the imperative for educational institutions to adopt appropriate measures to mitigate potential plagiarism issues, contributing to the ongoing debate surrounding the impact of AI technology on education. The research also delves into further implications arising from these findings.

**Khoury, R; Avila, AR; Brunelle, J & Camara, BM. 2023. How secure is code generated by ChatGPT? *arXiv preprint arXiv:2304.09655*.**

This paper explores the recent strides made by large language models in the realm of artificial intelligence (AI), with particular focus on ChatGPT—an AI chatbot developed and recently introduced by OpenAI, marking a significant advancement in the field. Beyond its proficiency in processing human-like text, ChatGPT demonstrates the unique ability to translate natural language into code. However, amidst these capabilities, the safety of programs generated by ChatGPT becomes a critical consideration. To address this concern, we conduct an experiment wherein ChatGPT is tasked with generating various programs, and we assess the security of the resulting source code. The study delves into whether ChatGPT can be influenced to enhance security through tailored prompts, and scrutinizes the ethical implications of leveraging AI for code generation. The findings suggest that while ChatGPT displays awareness of potential vulnerabilities, it often produces source code that is not sufficiently robust against certain attacks.

**Koubaa, A; Boulila, W; Ghouti, L; Alzahem, A & Latif, S. 2023. Exploring ChatGPT capabilities and limitations: a critical review of the NLP game changer.**

This paper delves into the recent groundbreaking natural language processing technology, ChatGPT, which emerged just three months ago, capturing significant attention for its exceptional capabilities. This AI milestone has prompted extensive scrutiny from researchers, industry professionals, decision-makers, and governmental bodies, keen on understanding the implications, threats, and benefits associated with this technology. Despite its relatively short existence, several researchers have already explored ChatGPT from diverse perspectives. Here, we offer a comprehensive review of ChatGPT, emphasizing its technical innovations in comparison to preceding models and scrutinizing existing research across various dimensions. Employing a meticulous methodology, we conduct a critical examination of the prevailing literature on ChatGPT, constructing a taxonomy to categorize diverse areas of study. Furthermore, we pinpoint future challenges and potential research trajectories linked to ChatGPT. Serving as the inaugural critical review of ChatGPT literature, this paper provides invaluable insights for practitioners and policymakers, establishing itself as a crucial reference for researchers aiming to advance studies on ChatGPT, spanning its applications and development.

**Koubaa, A; Boulila, W; Ghouti, L; Alzahem, A & Latif, S. 2023. Exploring ChatGPT capabilities and limitations: a survey. *IEEE Access*.**

This research investigates ChatGPT, an innovative natural language processing technology introduced a few months ago, which has garnered substantial attention for its remarkable capabilities. This milestone in AI has prompted scrutiny from researchers, industry professionals, decision-makers, and governments, delving into its implications, threats, and benefits. Despite its relatively brief existence, various researchers have analyzed ChatGPT from diverse perspectives. The paper offers a comprehensive review of ChatGPT, accentuating its technical advancements compared to prior models and scrutinizing existing research from multiple viewpoints. Employing a rigorous methodology, we conduct a critical examination of the existing ChatGPT research, establishing a taxonomy for various areas of study. Furthermore, we delineate future challenges and research trends associated with ChatGPT. This paper stands as the inaugural critical review of ChatGPT literature, providing valuable insights for practitioners and policymakers alike. It serves as a reference for researchers aiming to propel the research on ChatGPT, encompassing its applications and development.

**Koubaa, A; Qureshi, B; Ammar, A; Khan, Z; Boulila, W & Ghouti, L. 2023. Humans are still better than ChatGPT: case of the IEEEExtreme competition. *arXiv* preprint arXiv:2305.06934.**

This study takes a nuanced approach to the widespread acclaim of ChatGPT's outstanding performance by presenting a distinctive perspective. While previous research has emphasized ChatGPT's capabilities often comparable to or surpassing human skills in diverse tasks, this paper sheds light on a scenario where human proficiency outshines ChatGPT. The focus is specifically on the domain of computer programming, leveraging the IEEEExtreme Challenge competition—an esteemed annual international programming contest covering a spectrum of problems with varying complexities—as a benchmark. Through a meticulous evaluation involving a diverse set of 102 challenges from five IEEEExtreme editions and employing Python, Java, and C++, our empirical analysis reveals that human programmers excel in specific aspects of problem-solving within the programming context, contrary to prevailing notions. Notably, the average score attained by ChatGPT on the IEEEExtreme programming problems is 3.9 to 5.8 times lower than the average human score, depending on the programming language. This paper delves into these findings, offering critical insights into the limitations and potential avenues for enhancement for AI-based language models such as ChatGPT.

**Krause, D. 2023. ChatGPT and other AI models as a due diligence tool: benefits and limitations for private firm investment analysis. *SSRN* 4416159.**

This paper explores the increasing utilization of ChatGPT and other general AI models for investment due diligence, particularly in the context of evaluating private firms. Challenges arise in this domain as private companies, unlike their public counterparts, are not mandated to disclose extensive information, and the available data tends to be less standardized and comprehensive. Moreover, the complexity and opacity inherent in private markets further compound these challenges. To address these issues, machine learning techniques are employed as potential solutions. Sentiment analysis, for instance, can extract insights from unstructured data sources like SEC filings, social media posts, and news articles. This information, in turn, contributes to the construction of a comprehensive database comprising both structured and unstructured data for the evaluation of private firms. However, it is crucial to acknowledge the limitations of AI models in this context. These models may struggle to grasp the nuances of a private company's business model or its competitive landscape. As a prudent approach, it is emphasized that AI models should be used in conjunction with other due diligence methods, such as traditional financial analysis and industry research. This ensures a more holistic and nuanced evaluation of private companies, taking into account the multifaceted nature of their operations and market dynamics.

**Li, Z. 2023. The dark side of ChatGPT: legal and ethical challenges from stochastic parrots and hallucination. *arXiv* preprint arXiv:2304.14347.**

This paper underscores the transformative impact of Large Language Models (LLMs) like ChatGPT on our society, reshaping fundamental aspects of how we think, create, and exist. Notably, the integration of GPT in Bing has revolutionized our approach to online searching. Despite the numerous advantages offered by nascent LLMs, a new set of legal and ethical risks is surfacing, particularly concerning stochastic parrots and hallucination. While the European Union (EU) stands at the forefront of jurisdictions addressing AI model regulation, the emerging regulatory paradigm may underestimate the risks posed by these novel LLMs. Consequently, this communication highlights the necessity for further evolution in the European AI regulatory framework to effectively mitigate these risks.

**Liu, J; Xia, CS; Wang, Y & Zhang, L. 2023. Is your code generated by ChatGPT really correct? rigorous evaluation of large language models for code generation. *arXiv* preprint arXiv:2305.01210.**

This research delves into the long-standing study of program synthesis, with recent emphasis on leveraging the capabilities of Large Language Models (LLMs) for code generation. Traditional programming benchmarks, designed with curated synthesis problems and test cases, serve as a metric for evaluating LLMs' performance in code synthesis. However, these test cases may be limited in both quantity and quality, raising concerns about thoroughly assessing the functional correctness of the generated code in the era of LLMs. To address this concern, we introduce

EvalPlus, a code synthesis benchmarking framework that rigorously evaluates the functional correctness of LLM-synthesized code. EvalPlus enhances an evaluation dataset with an abundance of test cases generated automatically by a test input generator, utilizing both LLM- and mutation-based strategies. While EvalPlus is a general framework, we specifically augment the test cases of the well-known HUMANEVAL benchmark by 81 times to create HUMANEVAL+. Extensive evaluations across 19 popular LLMs, including GPT-4 and ChatGPT, demonstrate that HUMANEVAL+ effectively identifies previously undetected incorrect code synthesized by LLMs, reducing the pass@k by an average of 13.6-15.3%. This work not only reveals that previous evaluations of popular code synthesis may not accurately depict the true performance of LLMs but also suggests a new avenue for improving programming benchmarks through automated testing.

**Lo, CK. 2023. What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences*, 13(4): 410.**

This paper provides a rapid review of ChatGPT, an artificial intelligence-based chatbot launched in November 2022, exploring its capabilities across subject domains, applications in education, and concerns raised by researchers during its initial three months of release (December 2022 to February 2023). Utilizing a content analysis approach with 50 articles from relevant databases and Google Scholar, the review indicates varying performance in different domains, with strengths in areas like economics and programming but shortcomings in mathematics. While ChatGPT shows potential as an instructional assistant and virtual tutor, challenges include generating incorrect information and bypassing plagiarism detectors. The paper emphasizes the need for immediate updates to assessment methods and institutional policies, along with essential instructor training and student education to address ChatGPT's impact on the educational environment.

**Marron, L. 2023. Exploring the potential of ChatGPT 3.5 in higher education: benefits, limitations, and academic integrity. In: *Handbook of Research on Redesigning Teaching, Learning, and Assessment in the Digital Era*, pp 326-349. IGI Global.**

This paper investigates the potential advantages and challenges associated with the utilization of ChatGPT, a substantial language model, within higher education (HE), with a specific focus on its impact on academic integrity. Employing an exploratory research design, the study unfolds in three distinct phases. Phase 1 examines ChatGPT from the perspective of lecturers, Phase 2 from that of students, and Phase 3 through the lens of academic integrity. The primary objective is to assess whether ChatGPT can be employed to develop educational resources, facilitate collaborative projects, and support students in their academic endeavors. Additionally, the study explores the potential contribution of ChatGPT to instances of academic misconduct. It puts forth strategies to promote transparency and responsible AI use in higher education and provides recommendations for future research directions. In summary, the study asserts that AI can prove beneficial in higher education when employed responsibly and in tandem with effective pedagogical approaches.

**Memarian, B & Doleck, T. 2023. ChatGPT in education: methods, potentials and limitations. *Computers in Human Behavior: Artificial Humans*, 100022.**

This paper exams ChatGPT within the realm of public scrutiny, particularly in educational contexts. While public opinion has been a focal point, there is a notable gap in the analysis of studies conducted on ChatGPT in educational settings. In response, this review paper meticulously explores the diverse applications, potential benefits, challenges, and future directions of ChatGPT in education. A total of 63 publications were scrutinized using a general framework of open and axial coding. Through coding and summarization of methods, the review comprehensively addresses the potentials, limitations, and future work outlined in each study. Thematic analysis uncovers that existing studies in educational literature predominantly approach ChatGPT through a commentary and non-empirical lens. The potentials identified encompass personalized and intricate learning, specific teaching and learning activities, assessments, asynchronous communication, feedback, research accuracy, personas, and task delegation with cognitive offload.

Simultaneously, the review highlights various challenges that ChatGPT encounters or may face in an educational context, including plagiarism deception, potential misuse or lack of learning, issues of accountability, and privacy

concerns. Balancing both concerns and optimism, the paper emphasizes the critical need to safeguard student learning and uphold academic integrity in the integration of ChatGPT in education. This comprehensive review not only summarizes existing studies but also offers a unique and thorough discussion on future considerations for ChatGPT in the educational landscape.

**Mohammad, B; Supti, T; Alzubaidi, M; Shah, H; Alam, T; Shah, Z & Househ, M. 2023. The pros and cons of using ChatGPT in medical education: a scoping review. *Stud Health Technol Inform*, 305: 644-647.**

This paper conducts a scoping review to examine the merits and drawbacks of incorporating ChatGPT into medical education. A systematic search across databases including PubMed, Google Scholar, Medline, Scopus, and Science Direct identified 25 eligible studies out of 197 references. The study covers various applications of ChatGPT in medical education, encompassing automated scoring, teaching assistance, personalized learning, research support, quick information retrieval, case scenario and exam question generation, content creation, and language translation. Additionally, the paper addresses the challenges and limitations associated with ChatGPT in medical education, such as its inability to reason beyond existing knowledge, generation of inaccuracies, potential bias, impact on critical thinking skills, and ethical concerns related to exam and assignment integrity and patients' privacy.

**Morath, B; Chiriac, U; Jaszkowski, E; Deiß, C; Nürnberg, H; Hörth, K & Green, K. 2023. Performance and risks of ChatGPT used in drug information: an exploratory real-world analysis. *European Journal of Hospital Pharmacy*.**

This paper aims to assess the performance and risks associated with utilizing Chat Generative Pre-trained Transformer (ChatGPT) to address drug-related questions. A set of 50 drug-related questions was input into the ChatGPT software, and the resulting answers underwent standardized evaluation by six senior hospital pharmacists. The assessment criteria included content accuracy, patient management feasibility, and associated risks. Reference answers adhering to the German guideline of drug information were used as a benchmark. The findings revealed that only 13 out of 50 answers provided correct content and sufficient information to initiate management without patient harm. The majority of responses were either false (38%) or contained only partial correctness (36%), with no references provided. Approximately 26% of cases posed a high risk of patient harm, while 28% had a low risk. In high-risk cases, actions based on the provided information could potentially lead to patient harm. Furthermore, the study highlighted the lack of reproducibility in ChatGPT's answers over time, indicating substantial variability. In conclusion, the study underscores the challenges associated with the use of ChatGPT in drug information, citing issues such as incorrect content, absence of references, and limited reproducibility.

**Nguyen, XH; Nguyen, HA; Cao, L & Truong, H. 2023. Unleashing the potential and recognizing the limitations of ChatGPT. In: *Vietnamese Geography Education*.**

This paper studies the potential and limitations of ChatGPT, a cutting-edge language model, within the framework of Vietnamese geography education. The study seeks to analyze ChatGPT's capabilities in enhancing the teaching and learning of geography. By thoroughly exploring its features, including interactive engagement, personalized support, and access to extensive geographic information, the paper highlights the potential advantages of incorporating ChatGPT into geography classrooms. Concurrently, it critically evaluates the model's limitations, encompassing ethical considerations, potential biases, and challenges related to grasping complex geographical concepts. While ChatGPT exhibits potential in the context of Vietnamese geography education, it is imperative for the model to enhance its proficiency in geography-related tasks such as image processing and chart processing. This improvement is crucial for ChatGPT to become a more valuable tool for Vietnamese teachers and students in the realm of geography education. The findings derived from this analysis provide valuable insights, guiding the utilization of ChatGPT to maximize its advantages while conscientiously addressing its limitations, fostering a more effective and inclusive approach to geography education.

**Obaid, OI; Ali, AH & Yaseen, MG. 2023. Impact of Chat GPT on scientific research: opportunities, risks, limitations, and ethical Issues. *Iraqi Journal for Computer Science and Mathematics*, 4(4): 13-17.**



This paper explores the emergence of chatbots, specifically Generative Pre-trained Transformer (GPT), as a consequence of recent advancements in natural language processing (NLP). While Chat GPT has showcased considerable potential across various domains, including scientific research, its full impact is still unfolding. The aim of this paper is to delve into the possibilities, threats, limitations, and ethical considerations associated with Chat GPT in the realm of scientific research. The examination involves a comprehensive review of the existing literature on Chat GPT and its applications in scientific research, supplemented by illustrative case examples that highlight both the potential benefits and challenges in utilizing Chat GPT for scientific inquiry. The paper concludes by underscoring the ethical issues that necessitate attention before Chat GPT can be seamlessly integrated into the landscape of scientific research.

**Pegoraro, A; Kumari, K; Fereidooni, H & Sadeghi, AR. 2023. To ChatGPT, or not to ChatGPT: that is the question! *arXiv preprint arXiv:2304.01487*.**

This study addresses the global phenomenon of ChatGPT's widespread usage. As ChatGPT and other Large Language Models (LLMs) gain prominence, concerns regarding potential misuse, including the dissemination of fake news, plagiarism, manipulation of public opinion, cheating, and fraud, have escalated. Consequently, the imperative to distinguish between AI-generated and human-generated content becomes increasingly vital. Researchers have proposed diverse detection methodologies, ranging from basic binary classifiers to more advanced deep-learning models. These techniques vary in their reliance on statistical characteristics, syntactic patterns, semantic information, or contextual cues to enhance accuracy. The main objective of this study is to offer a comprehensive and up-to-date evaluation of the latest techniques in ChatGPT detection. Additionally, we assess the performance of other AI-generated text detection tools that do not explicitly claim to detect ChatGPT-generated content. To conduct our evaluation, we curated a benchmark dataset comprising prompts from ChatGPT and humans, encompassing diverse questions from medical, open Q&A, and finance domains, as well as user-generated responses from popular social networking platforms. This dataset serves as a benchmark for assessing the efficacy of various techniques in detecting ChatGPT-generated content. However, our evaluation results reveal that none of the existing methods can effectively detect ChatGPT-generated content.

**Poola, I. 2023. Overcoming ChatGPTs inaccuracies with Pre-Trained AI Prompt Engineering Sequencing Process. *International Journal of Technology and Emerging Sciences (ITJES)*, 3(3): 16-19.**

This research explores the role of artificial intelligence (AI) in advancing human inquiries through the lens of ChatGPT, a prominent language model. Developed by OpenAI, ChatGPT replicates human-like interaction, understanding context and generating relevant responses based on extensive training data. While it has gained acclaim for its capacity to effectively address diverse human inquiries with fluent and comprehensive responses, it is not without challenges and limitations. This study focuses on the difficulties and constraints encountered by ChatGPT in its usage. Addressing a variety of use cases, the research aims to illuminate the intricacies and limitations of the ChatGPT AI integration. The paper delves into the specific challenges faced by ChatGPT when employed with OpenAI, examining its shortcomings in logic, factual accuracy, mathematical understanding, coding, and potential biases. Additionally, it explores the broader social implications of ChatGPT's drawbacks. By categorizing these challenges into three distinct groups, the study provides valuable insights for researchers and developers aiming to enhance chatbots and language models. While large language models (LLMs) have demonstrated utility across various domains, this research uniquely focuses on uncovering and analyzing the specific limitations of ChatGPT, including hallucinations. The study underscores the ambiguity that arises in distinguishing text generated by ChatGPT from that authored by humans, given its ability to produce intricate essays, poems, functional code, and even web pages and charts based solely on text descriptions. Acknowledging ChatGPT's potential as a formidable competitor to popular search engines, the research contributes to a nuanced understanding of both the capabilities and limitations of this advanced language model.

**Qadir, J. 2023. Engineering education in the era of ChatGPT: Promise and pitfalls of generative AI for education. In: *2023 IEEE Global Engineering Education Conference (EDUCON)*, pp 1-9. IEEE.**



This paper delves into the dynamic landscape of engineering education, continually adapting to technological advancements and the evolving demands of the engineering industry. A notable development in this arena is the integration of generative artificial intelligence technology, exemplified by the ChatGPT conversational agent. ChatGPT holds the promise of delivering personalized and efficacious learning experiences by furnishing students with tailored feedback, explanations, and immersive virtual simulations for hands-on learning. However, a comprehensive exploration of this technology necessitates acknowledging its inherent limitations. Generative AI systems like ChatGPT are contingent on their training data and may inadvertently perpetuate biases or propagate misinformation. Moreover, the incorporation of generative AI in education raises ethical concerns, encompassing the potential for unethical or dishonest student use and the prospect of human unemployment due to technological redundancy. While the current state of generative AI technology, as exemplified by ChatGPT, showcases remarkable advancements with inherent flaws, it serves as a precursor to future developments. Engineering educators must grasp the implications of this technology and engage in a thorough examination of how to adapt the engineering education ecosystem. This adaptation aims to ensure that the upcoming generation of engineers can harness the benefits offered by generative AI while mitigating any adverse consequences.

**Rane, NL; Choudhary, SP; Tawde, A & Rane, J. 2023. ChatGPT is not capable of serving as an author: ethical concerns and challenges of large language models in education. *International Research Journal of Modernization in Engineering Technology and Science*, 5(10): 851-874.**

This paper delves into the dynamic role of ChatGPT and analogous large language models in the field of education, offering a comprehensive examination of their limitations, ethical concerns, and challenges. The research explores the implications of these models in various educational contexts, shaping the landscape of teaching, research, and scholarly communication. The exploration begins by scrutinizing the applications of ChatGPT in scientific writing and publishing. Critically assessing the constraints associated with its educational use, the paper acknowledges limitations in generating authoritative content, comprehending complex subjects, and ensuring information accuracy, presenting significant obstacles to integration into educational practices. Furthermore, the research addresses the ethical dilemmas and potential pitfalls arising from the extensive reliance on generative AI in education, touching on issues of bias, accountability, and the dissemination of misinformation. Emphasizing the importance of maintaining human agency and oversight, the paper underscores the need for responsible AI use in educational settings. It also investigates the impact of ChatGPT on academic research, examining its potential to enhance research productivity and the associated risks to the rigor and authenticity of scholarly work. Strategies and tools for detecting and mitigating academic misconduct involving AI-generated content are explored in detail. In addition, the research delves into the role of ChatGPT in fostering critical thinking skills among students, educators, and researchers. It explores innovative pedagogical methods that leverage generative AI to enhance critical thinking. The paper also considers the implications of ChatGPT on educational policy, addressing privacy concerns, intellectual property rights, and the need for regulations in the evolving landscape of AI in education. These insights are invaluable for educators, researchers, policymakers, and stakeholders aiming to harness the benefits of generative AI while navigating the associated challenges in the educational domain.

**Rao, H. 2023. Ethical and legal considerations behind the prevalence of ChatGPT: risks and regulations. *Frontiers in Computing and Intelligent Systems*, 4(1): 23-29.**

This paper provides a comprehensive overview of ChatGPT, encompassing its technical model, key capabilities, and various application scenarios. The analysis delves into ethical risks, including concerns about autonomy, moral conduct, human alienation, and value reconstruction. Additionally, legal risks are scrutinized, covering aspects such as legal entity status, copyright ownership, data security, and the opacity of algorithms. The paper then suggests regulatory principles, emphasizing tool positioning, safety, legality, transparency, credibility, fairness, and shared responsibility. Furthermore, it proposes a regulatory pathway that encourages the development of autonomous, safe, and controllable technologies, advocates for improvements in artificial intelligence legislation and justice, and underscores the need for establishing a diverse governance system.

**Ray, PP. 2023. ChatGPT: a comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*.**



This paper investigates the transformative impact of artificial intelligence (AI) and machine learning on the realm of scientific research, with a specific focus on the notable advancements in chatbot technology, exemplified by ChatGPT. Offering a comprehensive review, we delve into the background, applications, key challenges, and future directions of ChatGPT. The exploration begins by tracing its origins, development, and underlying technology, followed by an in-depth analysis of its diverse applications across industries such as customer service, healthcare, and education. Critical challenges faced by ChatGPT, including ethical considerations, data biases, and safety issues, are scrutinized, and potential mitigation strategies are discussed. The paper concludes by envisioning the future of ChatGPT, contemplating areas for further research and development. This includes exploring its integration with other technologies, enhancing human-AI interaction, and addressing concerns related to the digital divide. This review aims to provide valuable insights for researchers, developers, and stakeholders navigating the dynamic landscape of AI-driven conversational agents. Additionally, the study delves into the multifaceted ways in which ChatGPT revolutionizes scientific research, encompassing aspects such as data processing, hypothesis generation, collaboration, and public outreach. Ethical considerations and potential challenges associated with the use of ChatGPT in research are examined, emphasizing the need to strike a balance between AI-assisted innovation and human expertise. The paper also sheds light on biases and limitations inherent in ChatGPT, acknowledging the attention it has garnered across academia, research, and industries, despite controversies and ethical concerns.

**Rice, S; Winter, SR & Rice, C. 2023. The Advantages and Limitations of Using ChatGPT to Enhance Technological Research. SSRN 4416080.**

This research investigates ChatGPT, an innovative natural language processing technology introduced a few months ago, which has garnered substantial attention for its remarkable capabilities. This milestone in AI has prompted scrutiny from researchers, industry professionals, decision-makers, and governments, delving into its implications, threats, and benefits. Despite its relatively brief existence, various researchers have analyzed ChatGPT from diverse perspectives. The paper offers a comprehensive review of ChatGPT, accentuating its technical advancements compared to prior models and scrutinizing existing research from multiple viewpoints. Employing a rigorous methodology, we conduct a critical examination of the existing ChatGPT research, establishing a taxonomy for various areas of study. Furthermore, we delineate future challenges and research trends associated with ChatGPT. This paper stands as the inaugural critical review of ChatGPT literature, providing valuable insights for practitioners and policymakers alike. It serves as a reference for researchers aiming to propel the research on ChatGPT, encompassing its applications and development.

**Rimban, EL. 2023. Challenges and limitations of ChatGPT and other large language models. *International Journal of Arts and Humanities*, 4(1): 147-152.**

This paper delves into the challenges and constraints inherent in large language models, with a specific focus on ChatGPT as an illustrative example. Our exploration commences by examining the potential advantages offered by large language models, emphasizing their proficiency in generating natural language text and aiding in language-related tasks. However, we also acknowledge the prevailing concerns surrounding these models, encompassing environmental impact, potential biases, and the absence of interpretability. Moving forward, we scrutinize the specific challenges confronted by ChatGPT and similar models, highlighting limitations in contextual understanding, struggles with rare or out-of-vocabulary words, and a propensity to generate text that may be nonsensical or offensive. The paper concludes by putting forth recommendations for future research and development, underscoring the necessity for heightened transparency, interpretability, and ethical considerations throughout the creation and deployment phases of large language models.

**Rozado, D. 2023. The political biases of ChatGPT. *Social Sciences*, 12(3): 148.**

This study delves into recent advancements in Large Language Models (LLMs), signaling their imminent deployment in commercial applications as gateways for human interaction with technology and amassed knowledge. However, concerns about potential political biases embedded in these models raise issues of misuse. In this investigation, we present the findings of subjecting a state-of-the-art Large Language Model, the widely used

ChatGPT from OpenAI, to 15 different political orientation tests (14 in English, 1 in Spanish). The consistent results across these tests reveal that ChatGPT tends to be diagnosed with a preference for left-leaning viewpoints in 14 out of the 15 instruments. Interestingly, when directly questioned about its political preferences, ChatGPT often asserts a lack of political opinions, emphasizing its commitment to providing factual and neutral information. While it is desirable for public-facing artificial intelligence systems to furnish accurate and factual information on empirically verifiable issues, they should also strive for political neutrality, especially on normative questions lacking straightforward empirical validation. Ethical AI systems should aim to present users with balanced arguments, avoiding claims of neutrality when the content displays discernible signs of political bias.

**Sanchez-Ramos, L; Lin, L & Romero, R. 2023. Beware of references when using ChatGPT as a source of information to write scientific articles. *American Journal of Obstetrics & Gynecology*.**

This research aims to highlight a potential liability associated with the use of ChatGPT in the writing of scientific articles, particularly concerning references. While AI chatbots hold promise in streamlining the article-writing process, our investigation reveals a notable issue with ChatGPT. We have observed that ChatGPT frequently generates erroneous and fictitious references, encompassing incorrect authors, journal names, article titles, publication years, and PubMed identifiers (PMID and DOI). In some instances, the article title may be accurate, but the chatbot includes inaccurate information such as authors. Citations play a crucial role in attributing credit for contributions, identifying sources of ideas, and tracing the development of research. Errors in references not only undermine the credibility of authors but also compromise the trustworthiness of a scientific report. Consequently, it becomes imperative to verify all components of references provided by chatbots. The inaccuracies in ChatGPT's responses are attributed to the vast and diverse text data from various sources, coupled with inconsistencies or inaccuracies in the primary data. It is anticipated that as AI chatbot technology advances, these current limitations will likely see improvement over time.

**Sok, S & Heng, K. 2023. ChatGPT for education and research: a review of benefits and risks. *SSRN 4378735*.**

This paper explores the applications of Generative Pre-trained Transformer (ChatGPT), an artificial intelligence (AI) tool capable of rapidly generating detailed responses to prompts and subsequent queries. Introduced in November 2022 by OpenAI, an American AI research laboratory, using large language models, ChatGPT holds significant potential in education and research. The article delves into five key benefits of ChatGPT, including its role in creating learning assessments, enhancing pedagogical practices, providing virtual personal tutoring, aiding in outlining, and facilitating idea brainstorming. However, the discussion also addresses potential risks associated with academic integrity, issues of unfair learning assessments, dissemination of inaccurate information, and the risk of over-reliance on AI. The paper concludes by offering a set of recommendations to ensure the effective and responsible utilization of ChatGPT in educational and research contexts.

**Stahl, BC & Eke, D. 2024). The ethics of ChatGPT—Exploring the ethical issues of an emerging technology. *International Journal of Information Management*, 74: 102700.**

This paper delves into the ethical considerations arising from generative conversational AI systems, exemplified by ChatGPT. Employing established methodologies for examining the ethics of emerging technologies, the study systematically reviews potential benefits and concerns. By combining ethical issues from Anticipatory Technology Ethics, Ethical Impact Assessment, and Ethical Issues of Emerging ICT Applications with AI-specific concerns from the literature, the analysis evaluates ChatGPT's capacity to generate humanlike text and engage in seamless interactions. The findings suggest that while ChatGPT holds the potential for substantial societal and ethical benefits, it also raises noteworthy concerns pertaining to social justice, individual autonomy, cultural identity, and environmental impact. Key ethical concerns encompass responsibility, inclusion, social cohesion, autonomy, safety, bias, accountability, and environmental implications. In contrast to the prevailing discourse narrowly focusing on specific issues like authorship, this analysis systematically uncovers a broader and more balanced spectrum of ethical considerations that warrant attention. The results align with emerging research and industry priorities regarding the ethics of generative AI. Implications underscore the necessity for diverse stakeholder involvement, holistic consideration of benefits and risks during application development, and multi-level policy interventions to



foster positive outcomes. In essence, this analysis illustrates that applying established methodologies for technology ethics can establish a robust and comprehensive foundation to guide discussions and actions surrounding impactful emerging technologies such as ChatGPT, advocating for a sustained, broad ethical perspective as use cases unfold.

**Sun, X; Dong, L; Li, X; Wan, Z; Wang, S; Zhang, T & Wang, G. 2023. Pushing the limits of ChatGPT on NLP tasks. *arXiv preprint arXiv:2306.09719*.**

This research investigates the factors contributing to the subpar performance of ChatGPT on most Natural Language Processing (NLP) tasks compared to supervised baselines. The authors identify the limitations, including the token limit in prompts, the mismatch between ChatGPT's generation nature and NLP tasks, and intrinsic pitfalls of Large Language Models (LLMs) like hallucination and keyword overemphasis. The study proposes a set of general modules to address these issues and enhance ChatGPT's performance on NLP tasks. These modules include a one-input-multiple-prompts strategy, fine-tuned models for better demonstration retrieval, task transformation for improved alignment with the generation nature, reasoning strategies tailored to task-specific complexity, a self-verification strategy to address hallucination, and a paraphrase strategy for enhanced prediction robustness. The authors conduct experiments across 21 datasets covering representative NLP tasks, demonstrating that the proposed techniques significantly improve ChatGPT's performance, achieving results comparable to or surpassing supervised baselines and existing state-of-the-art performances.

**Tai, AMY; Meyer, M; Varidel, M; Prodan, A; Vogel, M; Iorfino, F & Krausz, RM. 2023. Exploring the potential and limitations of ChatGPT for academic peer-reviewed writing: addressing linguistic injustice and ethical concerns. *Journal of Academic Language and Learning*, 17(1): T16-T30.**

This paper examines ChatGPT, an OpenAI language model employing neural networks and the transformer architecture for Natural Language Processing (NLP) tasks. With its widespread popularity, reaching 100 million users in two months and receiving a multibillion-dollar investment from Microsoft, this commentary delves into the potential and limitations of utilizing ChatGPT for academic writing and publication. While it proves beneficial for editing tasks like spell and grammar checking, summarization, and translation, ethical concerns arise regarding the use of AI-generated text in academic contexts. ChatGPT's potential lies in addressing linguistic injustice for non-native English speakers in academic publishing, offering support to convey research findings more effectively. Writers can benefit from personalized feedback, improving writing style and gaining new perspectives. However, caution is advised, recognizing that ChatGPT's accuracy is constrained by the quality of input data, and it may generate incorrect text. Sole reliance on ChatGPT for writing assistance is not recommended.

**Tyson, J. 2023. Shortcomings of ChatGPT. *Journal of Chemical Education*, 100(8): 3098-3101.**

This paper addresses the notable gaps in the discourse surrounding the significant limitations of ChatGPT that hold relevance for chemical educators, students, and researchers. The examination of ChatGPT's performance, as detailed in three recent papers in *J. Chem. Educ.*, extends to a critical assessment of its capabilities in a 100-level general education chemistry course for nonscience majors. Additionally, the paper investigates ChatGPT's potential role as an assistant in literature searches related to the biogeochemistry of arsenic. A thorough analysis reveals that ChatGPT exhibits unreliability in performing mathematical operations, commits conceptual errors, and generates seemingly plausible citations with only partial accuracy in their purported contents. These shortcomings underscore the need for a comprehensive understanding of the limitations inherent in ChatGPT, particularly when considering its application within the domain of chemical education and research.

**Wach, K; Duong, CD; Ejdys, J; Kazlauskaitė, R; Korzynski, P; Mazurek, G & Ziemba, E. 2023. The dark side of generative artificial intelligence: a critical analysis of controversies and risks of ChatGPT. *Entrepreneurial Business and Economics Review*, 11(2): 7-24.**

This paper exams the challenges and limitations of incorporating generative artificial intelligence (GAI), with a specific focus on ChatGPT, within the business landscape. Employing a narrative and critical literature review, the study develops a conceptual framework, revealing the negative aspects of GAI development in management and



economics. Findings from an extensive review identify seven main threats, including the absence of AI market regulation, poor quality control, job losses due to automation, personal data violation, social manipulation, socio-economic inequalities, and AI technostress. Implications and recommendations emphasize the importance of AI market regulation, continuous education for adapting to job market shifts, prioritizing ethical considerations to address privacy concerns, and implementing responsible AI practices to mitigate risks. The contribution lies in drawing attention to the controversies and hazards associated with GAI, urging proactive measures for ethical deployment in business.

**Walters, WH & Wilder, EI. 2023. Fabrication and errors in the bibliographic citations generated by ChatGPT. *Scientific Reports*, 13(1): 14045.**

This paper assesses the limitations of chatbots, exemplified by ChatGPT, in the context of text generation and editing, particularly focusing on the issue of factually incorrect responses or hallucinations. The study concentrates on a specific type of hallucination, namely, the creation of bibliographic citations that do not correspond to legitimate scholarly works. Using both ChatGPT-3.5 and ChatGPT-4, the study generates literature reviews on 42 multidisciplinary topics, examining a total of 636 bibliographic citations across 84 papers. The evaluation involves scrutinizing the prevalence of fabricated citations, identifying errors in citations to non-fabricated papers, and assessing adherence to APA citation format. The findings reveal that, within this dataset, 55% of GPT-3.5 citations are fabricated compared to 18% of GPT-4 citations. Additionally, 43% of the real GPT-3.5 citations contain substantive errors, while this figure is reduced to 24% for GPT-4. Despite the improvements observed in GPT-4 compared to GPT-3.5, certain challenges persist.

**Wang, C; Liu, S; Yang, H; Guo, J; Wu, Y & Liu, J. 2023. Ethical considerations of using ChatGPT in health care. *Journal of Medical Internet Research*, 25, e48009.**

This paper explores the potential applications of ChatGPT in healthcare while emphasizing the importance of addressing ethical considerations to prevent possible harm. Ethical challenges related to legal, humanistic, algorithmic, and informational perspectives are identified. Legal ethics issues stem from the unclear assignment of responsibility in cases of patient harm and potential breaches of privacy. Clear legal guidelines are essential for liability allocation and user protection. Humanistic ethics concerns revolve around potential disruptions to the physician-patient relationship, humanistic care, and issues of integrity. Overreliance on AI may compromise compassion and trust, emphasizing the need for transparency and disclosure. Algorithmic ethics pose questions about bias, responsibility, transparency, explainability, validation and evaluation. Information ethics encompass concerns about data bias, validity, and effectiveness, where biased training data can lead to biased output, impacting patient adherence and encouraging self-diagnosis. Ensuring accuracy, reliability, and validity of ChatGPT-generated content necessitates rigorous validation and ongoing updates aligned with clinical practice. Navigating the evolving AI ethical landscape in healthcare requires adherence to strict standards. Comprehensive ethical guidelines empower healthcare professionals to responsibly use ChatGPT, facilitate accurate information exchange, safeguard patient privacy, and enable informed healthcare decisions.

**Wu, X; Duan, R & Ni, J. 2023. Unveiling security, privacy, and ethical concerns of ChatGPT. *Journal of Information and Intelligence*. <https://doi.org/10.1016/j.jiixd.2023.10.007>**

This paper explores the landscape of ChatGPT, an AI-powered chatbot employing topic modeling and reinforcement learning for natural response generation. While ChatGPT presents considerable potential in diverse sectors, including customer service, education, mental health, personal productivity, and content creation, addressing its security, privacy, and ethical implications is imperative. By tracing the evolution from GPT-1 to GPT-4 and examining the model's features, limitations, and potential applications, this study aims to unveil the potential risks associated with integrating ChatGPT into everyday scenarios. With a focus on security, privacy, and ethical concerns, the paper outlines the challenges these issues pose for widespread adoption. Lastly, the study analyzes the outstanding problems in these domains, advocating for collaborative efforts to ensure the development of secure and ethically sound large language models.



**Yu, Y. 2023. Discussion on the Reform of Higher Legal Education in China Based on the Application and Limitation of Artificial Intelligence in Law Represented by ChatGPT. *Journal of Education, Humanities and Social Sciences*, 14: 220-228.**

This paper examines ChatGPT as a representative of generative artificial intelligence, representing a disruptive innovation in the AI field. With its formidable natural language processing capabilities, ChatGPT has found widespread utility in the legal profession, offering robust support to the industry. As ChatGPT and other AI technologies undergo continuous evolution and broader adoption, their inevitable influence on the legal sector and current higher education in Chinese law will be profound. In light of this, the paper delves into the challenges and opportunities that Chinese legal education faces in the era of artificial intelligence. The focus centers on exploring the applications and limitations of generative AI, such as ChatGPT, within the legal profession.

**Zhuo, TY; Huang, Y; Chen, C & Xing, Z. 2023. Red teaming ChatGPT via Jailbreaking: bias, robustness, reliability and toxicity. *arXiv preprint arXiv:2301.12867*.**

This study delves into recent advancements in natural language processing (NLP), exploring the synthesis and comprehension of coherent text in an open-ended manner, translating theoretical algorithms into practical applications. Large language models (LLMs) have significantly impacted various sectors, including report summarization software and copywriting. However, concerns have arisen regarding the potential social prejudice and toxicity exhibited by LLMs, raising ethical and societal risks. The need for large-scale benchmarks for accountable LLMs becomes apparent. While some empirical investigations highlight ethical challenges in advanced LLMs, a systematic examination and user study of the risks and harmful behaviors associated with current LLM usage are lacking. To contribute to responsible LLM development, we employ a qualitative research method called "red teaming" on OpenAI's ChatGPT1, focusing on understanding practical ethical dangers in recent LLMs. The analysis covers four perspectives: Bias, Reliability, Robustness, and Toxicity. We comprehensively benchmark ChatGPT on multiple sample datasets aligned with these perspectives. Our findings reveal that existing benchmarks may not address numerous ethical risks, which we illustrate through additional case studies. We also discuss the implications of our discoveries on AI ethics and harmful behaviors in ChatGPT, along with future challenges and considerations for the responsible design of LLMs. This study aims to shed light on future efforts to identify and mitigate ethical hazards associated with LLM applications.

**Author contributions** S He proposed and contributed to the whole research, interpretation, and writing of the paper. The author prepared and reviewed the manuscript and approved the final version of the manuscript.

**Data availability statement** The data that support the findings of this study is available from the author upon reasonable request.

**Declarations** The author declare that he has no conflict of interest.

