Information Sciences with Applications, Vol. 3, 2024
https://doi.org/10.61356/j.iswa.2024.3303

A Critical Evaluation of ChatGPT's Adherence to Responsible Artificial Intelligence Principles

Sami Lababidi [1,*]

[1] School of Engineering and Applied Science, University of Colorado, Boulder, CO 80309, USA; Sami.lababidi@colorado.edu.

* Correspondence: Sami.lababidi@colorado.edu.

Abstract: The swift evolution of ChatGPT maintains revealing great promise in different fields of life while occasionally with ethically questionable impacts. While the current research effort has focused on the benefits that can be gained from ChatGPT, increasing concerns have been raised about the ethical implications that could result from its widespread use. To this end, this study presents an in-depth investigation of the ethical aspects of ChatGPT from the perspective of responsible AI. In particular, a novel theoretical framework is introduced to practically analyze and interpret the ChatGPT from an ethical side lens. Our framework is based on the concept of responsible AI, to focus on the variety of scenarios in which ChatGPT can possibly lead to unintentional consequences, and to advocate alternate paths that the researcher and practitioners can follow to expand their knowledge regarding the mitigation of such incidences. This work expands the theorization of the ethical side to disclose unknown ideas of existing literature and to suggest other leading premises that may guide future development and use of ChatGPT and alike language models.

Keywords: Large Language Models, ChatGPT, Responsible AI, Ethics, Robustness, Accountability, Bias.

1. Introduction

Natural language processing (NLP) advances have shown promise in data-rich fields and in enhancing people's daily lives in general. Web search, language modification, and content production are just a few of the many real-world applications that make use of linguistic models. Yet, users may feel disconnected from the app's content due to a lack of connection and involvement [1]. The impact of language model robots like Amazon Echo and Google Home on people's daily lives could be enormous due to the universality of natural language in human communication. When compared to other linguistic AI tools, OpenAI's proposed ChatGPT stands out for its numerous useful features that make it a robust language model with broad potential applications. It's intelligent enough to follow the thread of a conversation and join in where necessary [2]. It can generate responses in a range of tones (official, informal, and funny) and in a number of languages. It can do things like write music and write computer programmes as well as finish assignments. ChatGPT is powered by a sophisticated pre-trained language model to rapidly grasp user enquiries and generate answers that sound natural. The widespread influence and active participation of big communities in ChatGPT's role as a helpful tool contribute to its rising profile as a well-known question-answering system, which has been reported on in the usual global media [3]. After being trained on a large corpus of text, ChatGPT may generate novel textual answers based on the information it has learned. Typical users may have trouble distinguishing between human and automated writing in many contexts. It's not always the case, though, especially among experts and academics [4]. Without any training, it can write functional code, create charts and webpages from textual descriptions, and make sophisticated literary works. In light of its spectacular responses, ChatGPT has proven itself to be a formidable rival to the industry standard search engine [5].

While previous studies on huge language models have focused on the potential benefits of the field's discoveries, it's interesting that no one has yet used a "ethical side" analogy to investigate the difficulties that have been overlooked or hidden. The purpose of this research is to bring up potential issues with ChatGPT and to encourage a critical analysis of those issues. Difficulties that go unreported due to their being deemed weird or deviant are rarely subjected to in-depth investigation. However, by dissecting these types of problems, we may enhance the state-of-the-art large language models (such as ChatGPT) in a trustworthy, efficient, and principled fashion. In light of the fact that they may gain insight from unwanted or unintended repercussions of AI and work proactively towards their growth, responsible AI pillars are best paired with a holistic view of AI through an ethical-side lens. A more complete and nuanced understanding of the ChatGPT is made possible by the ethical-side analysis, which allows us to consider settings or situations that have been missed by traditional research. The ethical implications of AI are controversial. The design, development, and validation of AI technologies and systems must take into account ethical and legal considerations, especially those related to human values. The second is to use artificial intelligence methods to achieve the objectives of Responsible AI.

Recently, AI has shown its usefulness to society in numerous ways, some of which raise ethical concerns. Consequences for humanity are mounting as a result of irresponsible AI development in areas as diverse as data security, privacy, and bias, as well as in areas of human rights and environmental fairness. The key to long-lasting innovation has been highlighted as creating and retaining public trust in AI. As a result, there is a lot of focus on the question of how to build AI that is ethical or responsible. Companies, universities, and government agencies have all developed their own sets of ethical AI standards and principles, with about a hundred in all. Nevertheless, rules and regulations alone won't do much to guarantee the reliability of AI systems. The situation is further complicated by the fact that different people and different cultures have different ideas about what constitutes confidence in AI, and these ideas don't always line up with how trustworthy any given AI system actually is. Despite research efforts devoted to the study of the ethical sides of the language model, there are significant research gaps that are summarized as follows. First, the literature on responsible AI is concentrated on general and theoretical discussions of AI solutions, which might not correctly indicate the practical ethical consequences, especially in the case of ChatGPT. Second, due to the rapid development of the NLP field, there has not been enough time spent on an ethical analysis of new language models. Third, Individuals on a regular basis do not agree on the ethical dangers posed by the present applications of improved language models. Fourth, the majority of investigations only attempt to quantify a subset of ethical concerns, rather than addressing all of them head-on.

In response to the above challenges, this study provides a provide holistic study of how ChatGPT obeys or violates the principles of responsible AI. This paper presents a holistic qualitative investigation and record of responsibility dimensions in ChatGPT, as a use case for practical language. In particular, we introduce a novel theoretical framework to drive insightful analysis of the ChatGPT from the perspective of responsible AI principles. Our framework has demonstrated several conducts exposed by ChatGPT that are likely to violate the principles of responsibility such as fairness, transparency, precision, robustness, and accountability. In the course of our work, we hope to get closer to having a comprehensive comprehension by making use of our existing frameworks, intuition, and instances, in addition to the evaluations provided by humans. We have high hopes that this research will yield important insights that will help promote future work on recognizing and reducing the ethical risks posed by language models and the applications of those models.

The remaining part of this paper is structured as follows. Section 2 provides an overview of the literature studies on the ethical aspects of language models. Section 3 background on responsible AI in the context of language generation. Section 4 provides an in-depth analysis of ChatGPT's impacts on responsible AI. The main conclusions are drawn in section 5.

2. Literature

With the rapid adoption of AI-generated texts in many fields, the research community is encountering many questions regarding the ethicality of AI-generated texts. In line with this growing interest, a variety of studies have investigated the need for policies to regulate the use of AI-generated texts. For example, the authors of [1] experimentally investigated the ability of people to differentiate and favor AI-generated text versus human-written one by performing two experiments to evaluate the behavioral consequences of GPT-2. In their experiments, the early lines of randomly selected human poems were used to make GPT-2 generate some poems. The findings show that the participants are unable to dependably identify the AI-generated poems in Human-in-the-loop settings, but they identify them successfully in Human-out-the-loop scenarios. In [2], the authors performed three experiments to identify if the AI-generated texts are plausible and could impact thoughts on unknown policy. In the first experiment, the human's ability to perceive the AI-generated text is evaluated with respect to the initial story, while the second one examined the collaboration between devotion and AI-created news. In the third experiment, the authors examined the distributions of observed integrity across various volumes of the AI model. From these experiments, the authors found that people are principally unable to differentiate between AI- and human-generated text. In [3], The scholars tried to study the possible political and social threats evolving through the malevolent usage of text generation tools, and their possible influence on human civil rights. They conducted an experiment to drive observations in the elegance of political directors through refinement of the pre-trained AWD- LSTM model on dialogues data prepared by UN General Assembly. In [4], the authors examined how human writers work together with ML models in fantasy writing, by prototyping a human-AI collaborative writing software that enables human authors to condense, edit, summarize, and revive AI-generated text, then used "finish each other's story" strategy to explore the dynamics of human-AI collaboration. In [5], the authorship and ownership of human-AI writers are studied for the personalized LLMs. They demonstrated AI ghostwriter influence, in which the users did not regard themselves as the authors or owners of AI-generated text but desisted from openly proclaiming AI authorship. The study also discussed that personalization had no influence on the AI ghostwriter influence and regulating the model boosting users' perception of ownership. In [6], the authors studied the role of AI writing in scientific content by developing a feature narrative approach to differentiate between human-written texts and AI-generated ones according to syntactical, grammatical, and semantical evaluations. Multiple approaches were explored to the gap between human-written and AI-generated scientific text by AI-generated scientific text detection models. In [7] the authors studied the ethical problems in ChatGPT as an example of practical LLM. In their work, they qualitatively researched the practical features of ChatGPT to analyze their ethical threats from four distinct viewpoints, namely bias, reliability, robustness, and toxicity. In addition, the authors of [8] provided an extensive failure analysis of ChatGPT making use of information sourced from Twitter, which resulted in eleven areas of failures. The work [9] studied and evaluated the robustness of the ChatGPT against adversarial and out-of-the-distribution samples for various linguistic tasks. The authors of [10] developed a generic model for evaluating LLMs to evaluate human characters according to Myers Briggs Type Indicator (MBTI) examinations. In their work, they devised unbiased urges by unsystematically transposing preferences in MBTI inquiries and leveraged the means of testing results to promote a more unbiased generation of answers. In [11], lecturers were interviewed and questioned to share their practices to be analyzed to obtain an in-depth interpretation of how ChatGPT impacts learning. The authors claimed that ChatGPT can provide helpful information that facilitated interpreting the learning materials. However, this information still needs confirmation from the original sources. Beyond the above studies, the literature is rapidly evolving in exploring the potential use case of ChatGPT, for instance, in [12], the authors tried to investigate the potential security threats of ChatGPT. Moreover, in [13], and [14], the authors tried to discuss the role of AI in global warming and healthcare domains,

respectively. However, the lack of investigation of the ChatGPT and similar language models from an ethical-side lens.

3. Preliminaries and Foundations

The exponential growth of artificial intelligence (AI) technologies and the profound changes they are bringing about in our lives is without precedent. Concerns about the veracity and responsibility of AI systems have arisen in response to the consequences of AI in our lives. Some of these concerns are listed below:

- If an error occurs within the AI system, who or what is the system component to blame?
- Why does an AI system recommend those choices or make those predictions?
- How can we guarantee that computer-generated verdicts and predictions are made in a humane and accountable fashion?

In the fact that they don't rely on random associations or include hidden biases. The research attention moved toward techniques for modeling and reasoning about responsibility, which can contribute to addressing such concerns and thereby guarantee the dependability of AI systems. In previous studies, the concept of responsibility, and associated terms like accountability and liability, are leveraged in different manners with a variety of connotations. For instance, determining which component of an AI system (e.g., ChatGPT) is to blame for a certain result is distinct from determining how to sustainably utilize AI tools (e.g., ChatGPT). The later questions concentrate on how different stakeholders, such as the developers or the end consumers of AI systems, consider societal, legitimate, and ethical standards like privacy regulations, societal fairness, and subjectivity, while the first pertains to causation and capacity of the involved materials, including human and artificial systems. The earlier understanding of responsibility is based on a more basic understanding of causation, which might help to explain how responsibility is used in different ways throughout the research on AI. It’s possible to wonder whether specific AI systems or components of those systems are responsible for a certain result. Concerns like whether particular data-driven judgments or forecasts are driven by incidental relationships in the data and if the data is devoid of unwanted or unexpected biases may be of importance in other contexts. In the former, we might care about specific events that lead to a certain result, whereas in the latter, we might want to know how various occurrences in a community are related to one another. The above scenarios motivated the representatives, research community, and developer to think about in what way the development and application of AI tools (such as ChatGPT) can follow the standards of responsible AI. As a result, AI creates new ethical, legislative, and organizational problems, such as unintentional discriminatory practices, skewed results, and problems with users' awareness of and knowledge of how AI affects decisions. While earlier research focused on isolated aspects of responsible standards, such as eliminating prejudice, making AI results understandable, or security and privacy, recent years have seen a shift towards a more comprehensive view of what responsible AI entails.

Developing an agreement on the importance of ethical, accessible, and responsible usage of AI technology to meet customers' requirements, organizational values, and societal standards underpins the concept of responsible AI. The concept of responsible AI encompasses a wide range of needs that must be addressed throughout a system's development and maintenance. Some governments have defined the basic principles that underpin responsible AI and those businesses, both commercial and public, should be required to follow, demonstrating the critical significance of responsible AI at the legislative level. For example, a new sub-index has been added to the AI readiness index, which quantifies the extent to which responsible AI concepts are accepted and used in the implementation of AI technology. Microsoft Corporation reported six principles for responsible AI including Fairness, Reliability & Safety, Reliability & Safety, Privacy & Security, Inclusiveness, Transparency, and Accountability. On the other hand, Facebook released a slightly different set of principles namely Privacy & Security, Fairness & Inclusion, Robustness & Safety, Transparency & Control, and Accountability & Governance. Google also defined its principles that make required the AI to be 1)

communally beneficial; 2) avert unfair bias; 3) Be safe; 4) Be accountable; 5) preserve privacy; 6) obey high-level standards for scientific excellence; 7) Be used according to these ethics. More institutions have researched the principles of responsible AI from different standpoints[1],[2] (See Figure 1).

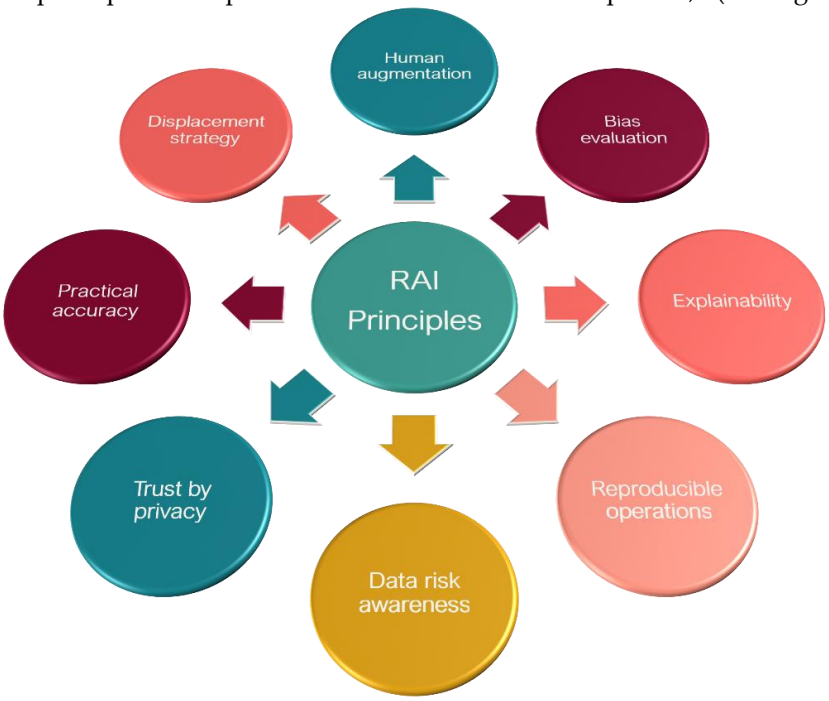

Figure 1. Illustration of the eight principles of responsible AI.

Empirical research at the institutional, business entity and personal levels is usually scarce, despite the growing body of literature on the topic of responsible governance of AI applications at the social and regulatory levels. Despite the dearth of empirical research in these settings, the conventional research collection has indeed been capable of recognizing and defining the primary aspects that constitute responsible AI. This literature has been generated from the viewpoints of a wide range of stakeholders, including businesses, universities, consultancies, institutions, non-profits, and even multilateral discussion boards and associations. Several facets of responsible AI are interdependent, and it is suggested that they are essential for reducing the likelihood of undesirable outcomes brought about by AI.

4. Theoretical Framework for Analyzing the Responsibility Principles of ChatGPT

In this section, we propose a theoretical framework for developing a group of ideas, theories, and analytics that enable a holistic interpretation of the responsibility dimension of the evolving ChatGPT (See Figure 2). The essential objective of our framework is to deliver insights to determine a firm foundation for the theoretical perspective that guides the community in selecting the research design for language models, training of language models, ethical analysis of language models, and explanation of their results. This, in turn, will help the researcher to develop a clear interpretation of the ChatGPT-alike language models by identifying their research gaps, developing hypotheses, and making predictions about their outcomes, in a way that enables the researcher community to connect the findings with the existing body of knowledge in responsible AI. In the following subsections, we dive into the details of each ethical dimension of ChatGPT in our framework.

[1] https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles
[2] https://ethical.institute/principles.html#commitment-1

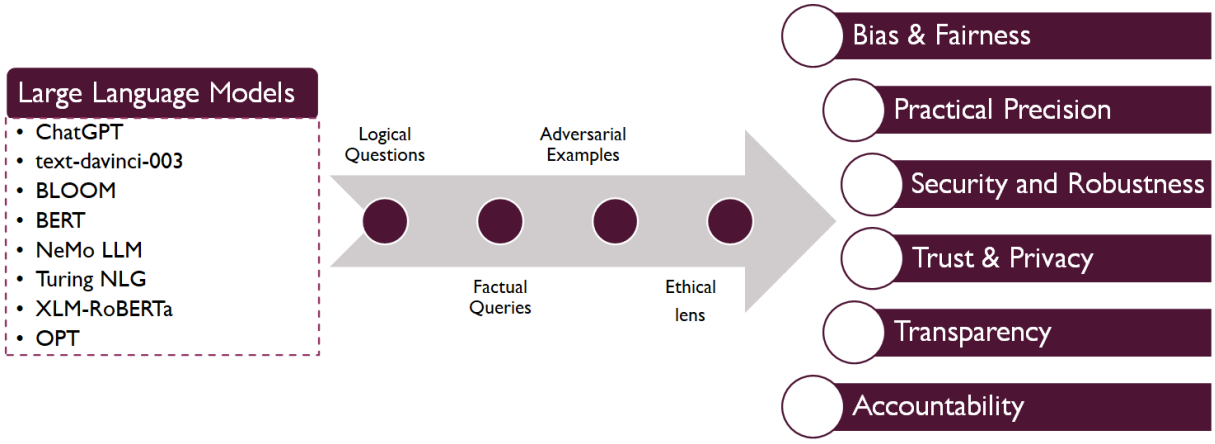

Figure 2. Illustration of the architecture of our proposed theoretic framework for analyzing the responsibility aspects of the ChatGPT.

4.1 Bias Problem in ChatGPT

Algorithmic bias is another name for ML bias. Systematically biased findings are the result of erroneous presumptions made during the machine learning process or biases present in the training data. Discrimination based on gender, race, age, and other characteristics has been documented in language models. Human prejudice, whether consciously or unconsciously, manifests itself in ML's biased solutions. Here, a cognitive bias is a reflection of how one feels about a particular individual or community because of their assumed membership in that group. Researchers have identified and categorized numerous cognitive prejudices that have a negative impact on the efficacy of language models.

4.2 Security and Robustness

The security principle emphasizes developing acceptable practices and infrastructure that safeguard the security of both data and language models during the development and testing phases. Autonomous language generation systems paved the way for new possible security violations. Given the growing popularity of ChatGPT across the globe and in a broad variety of applications, it is urgently important to assess the risks associated with it. To our knowledge, only [9] has evaluated the robustness of chatGPT, while previous studies have examined multiple facets of chatGPT in governance, morals, education, and logic. The ability to operate normally and produce accurate results despite interference from outside sources is what we mean when we talk about a system being robust. It has significant value in real-world uses, particularly in safety-sensitive ones. For example, if ChatGPT is applied to identify false information, a mischievous user may include specific perturbations into the contents to fool the system to generate incorrect decisions. The dependability of the language model can significantly degrade if it was not robust against such an adversary. In the study [9], the authors investigated the zero-shot robustness of ChatGPT against adversarial samples as well as out-of-distribution samples. The former emphasized the case of altering individual words or phrases, adding noise or synonyms, or modifying the order of words in the sentence to fool the language model. The latter case emphasized evaluating the ChatGPT on unseen samples drawn from a new distribution that varies from the distribution of training data. Following the same ideology, our theoretical framework analyzes the robustness of ChatGPT against sentence-level, word-level, char-level, and multi-level adversarial samples, as well as out-of-distribution samples. Conforming to the [9], our framework the ChatGPT, and competing language models are away from precision in either classification of translation tasks. It cannot also drive ultimate answers for medical-related problems, but rather, it provides advised recommendations and analysis.

4.3 Practical Precision of ChatGPT

As a pillar of responsibility, practical precision imposes a commitment to developing procedures to guarantee high-precision functionality that is better aligned with the underlying tasks. When developing large language models, such as ChatGPT, that learn from multi-domain data, it is critical to acquire a comprehensive interpretation of the underlying measures to evaluate their performance. For this reason, it is not always sufficient to rely solely on default/basic cost indicators or measures of precision, as what is "correct" for a machine may be incorrect for a person (and vice-versa). The precision of ChatGPT can be practically improved by dissecting the consequences of performance measures from a domain-dependent viewpoint and exploring different cost functions informed by domain knowledge. Inappropriate evaluation of ChatGPT can lead to many failure cases, in which misleading and incorrect answers can be generated. In [8], the authors provided a holistic analysis of the ability of ChatGPT from a failure standpoint, whereby the case of the failure was grouped into eleven categories. However, most of these failures no longer occur in the current editions of ChatGPT since the technology is encountering a rapid evolution. In this regard, our framework introduces and analyzes some instances of ChatGPT failures under the collection of randomly selected reasoning problems. However, we did not present any taxonomy for categorizing the failures since it is beyond the scope of our framework.

4.4 Transparency of ChatGPT

The term "transparency" principle appears frequently in current discussions of responsible AI, especially in regard to facilitating greater clarity on AI solutions and the processes surrounding responsibility and liability [15]-[17]. One of the dangers of ChatGPT, which is highlighted by our framework, is that it can make choices less clear and further remove them from the involvement of a human agent, thereby reducing accountability. Because of these occurrences, "explainable AI" has become a necessity, meaning that the results and data utilized to draw a decision from ChatGPT, must be documented in a manner that may be understood by various levels of users. However, Explainable AI, faces several challenges, such as the difficulty of systematically transferring the method by which complicated algorithmic language models produce results and the difficulty of adapting these models to new information in real-time. Simultaneously, there is a long list of factors like the sort of in-hand ChatGPT result, the context, timing, and the importance of the AI judgment that must be considered when discussing explainable AI at the user level. Additionally, human interpretation is introduced into the process whenever computer output is communicated directly to a human user. As a result, studying how people react to various AI interpretations and how well they meet their needs is an exciting field for future study. Nevertheless, increasing the explainability of language model might lead to less practical precision and vice versa. This way, the explainability of ChatGPT develops a double-edged weapon: While greater transparency would allow for easier verification and appeals of decisions, it would additionally raise the likelihood of mistakes being made. Therefore, greater explainability might not be viewed as beneficial by the individual who will be negatively impacted by that decision. Up till now, the literature has not studied the trade-off between practical precision and explainability of decisions made by language models (including ChatGPT). Despite the incorrect answers given by the ChatGPT provide explanation or justification of its results. This explanation can be regarded as kind of fake or non-reasonable explanation that is not supportive to transparency principle at all.

Transparency in AI encompasses more than just explainable AI; it also includes tracking and dialogue [18]. This presents an important gap for scholars working to define ChatGPT governance practices that designate anchors at various stages of AI projects to pinpoint the causes of wrong ChatGPT choices. The need for audibility is also met by allowing for such documentation at all stages of the procedure. Therefore, it is important to investigate how distinct demands on traceability influence the development of AI governance practices in various settings.

4.5 Accountability

Accountability is a foundation stone of the authority of AI, which is usually demarcated too roughly for the reason that its complicated nature and the socio-technical construction of language models suggest a diversity of principles, applies, and metrics to which accountability in language models may refer [19-20]. The documented unethical side effects of AI accountability highlight the importance of principles influencing the development and implementation of ChatGPT. By providing this framework, we hope to persuade managers that the positive, explicable, and clear results of the ChatGPT are worth avoiding. Calling attention to the need for ChatGPT's underlying procedure measures in the design, implementation, and accountability of organisations [21]. There is a need for more study into how accountability as a critical principle fueling large language models affects the creation, assessment, and use cases, which is demonstrated here by the significance and some choices to enhance the accountability of ChatGPT. For instance, knowing who to hold liable for unintended consequences is critical for the success of any large-scale language modelling effort. Examining the effects of accountability on language model deployments requires an appreciation of this difference and the identification of how accountability modifies work practices and the distribution of responsibilities in the digital world. It also emphasizes the importance of standardizing best practices to make auditable the development procedure for ChatGPT and similar tools. Finally, the issue of accountability is intrinsically linked to the kinds of testing done to guarantee that ChatGPT does not cause any unethical consequences on the users in either a direct or indirect manner. The issue of how to ensure, at various stages of development, that ChatGPT meets a necessary threshold of trust, and how that ought to be evaluated, is one of the most pressing in the field of language model development and testing [22].

5. Conclusions

This work provides a critical overview of the ChatGPT-generated text to problematize the field and believe according to an ethical perspective on the implications of using ChatGPT in various applications. A theoretical framework is presented to leverage the ethical-side lens to enlighten a more nuanced interpretation of how ChatGPT is employed in reality, its adverse or unintentional implications, and their justification of them. This way, we provide the community with scenarios, in which ChatGPT fails and proactively distinguish scenarios in which ChatGPT violates the principles of responsible AI. Our framework is built on responsible AI to formulate a collection of research queries regarding the potential use and ethical aspects of the ChatGPT. While our framework is not comprehensive, it does provide a foundation for more comprehensive research on how to approach the ethical use of ChatGPT by highlighting the significance of using the ethical-side lens for speculating and problematizing the evolving large language models. The findings from the ethical-side lens serve demonstrated that while the principles of responsible AI are important in developing trustworthy large language models, they are also relatively abstract and high-level and can't deliver much supervision for practitioners concerning the developments and practical use of large language models.

Funding

This research was conducted without external funding support.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Conflicts of Interest

The author declares that there is no conflict of interest in the research

.Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

Not applicable

References

[1] Köbis, N., & Mossink, L. D. (2021). Artificial intelligence versus Maya Angelou: Experimental evidence that people cannot differentiate AI-generated from human-written poetry. Computers in human behavior, 114, 106553.

[2] Kreps, S., McCain, R. M., & Brundage, M. (2022). All the news that's fit to fabricate: AI-generated text as a tool of media misinformation. Journal of experimental political science, 9(1), 104-117.

[3] Bullock, J., & Luengo-Oroz, M. (2019). Automated speech generation from UN General Assembly statements: Mapping risks in AI generated texts. arXiv preprint arXiv:1906.01946.

[4] Yang, D., Zhou, Y., Zhang, Z., Li, T. J. J., & LC, R. (2022, March). AI as an Active Writer: Interaction strategies with generated text in human-AI collaborative fiction writing. In Joint Proceedings of the ACM IUI Workshops (Vol. 10).

[5] Draxler, F., Werner, A., Lehmann, F., Hoppe, M., Schmidt, A., Buschek, D., & Welsch, R. (2023). The AI Ghostwriter Effect: Users Do Not Perceive Ownership of AI-Generated Text But Self-Declare as Authors. arXiv preprint arXiv:2303.03283.

[6] Ma, Y., Liu, J., & Yi, F. (2023). Is This Abstract Generated by AI? A Research for the Gap between AI-generated Scientific Text and Human-written Scientific Text. arXiv preprint arXiv:2301.10416.

[7] Zhuo, T. Y., Huang, Y., Chen, C., & Xing, Z. (2023). Exploring ai ethics of chatgpt: A diagnostic analysis. arXiv preprint arXiv:2301.12867.

[8] Borji, A. (2023). A categorical archive of ChatGPT failures. arXiv preprint arXiv:2302.03494.

[9] Wang, J., Hu, X., Hou, W., Chen, H., Zheng, R., Wang, Y., ... & Xie, X. (2023). On the Robustness of ChatGPT: An Adversarial and Out-of-distribution Perspective. arXiv preprint arXiv:2302.12095.

[10] Rao, H., Leung, C., & Miao, C. (2023). Can ChatGPT Assess Human Personalities? A General Evaluation Framework. arXiv preprint arXiv:2303.01248.

[11] Firaina, R., & Sulisworo, D. (2023). Exploring the Usage of ChatGPT in Higher Education: Frequency and Impact on Productivity. Buletin Edukasi Indonesia, 2(01), 67-74.

[12] Sebastian, G. (2023). Do ChatGPT and other AI chatbots pose a Cybersecurity risk?-An Exploratory study. An Exploratory Study (February 19, 2023).

[13] Biswas, S. S. (2023). Potential Use of Chat GPT in Global Warming. Annals of Biomedical Engineering, 1-2.

[14] Biswas, S. S. (2023). Role of Chat GPT in Public Health. Annals of Biomedical Engineering, 1-2.

[15] Mitrović, S., Andreoletti, D., & Ayoub, O. (2023). Chatgpt or human? detect and explain. explaining decisions of machine learning model for detecting short chatgpt-generated text. arXiv preprint arXiv:2301.13852.

[16] Kiseleva, A., Kotzinos, D., & De Hert, P. (2022). Transparency of AI in healthcare as a multilayered system of accountabilities: Between legal requirements and technical limitations. Frontiers in Artificial Intelligence, 5, 82.

[17] Andrada, G., Clowes, R. W., & Smart, P. R. (2022). Varieties of transparency: Exploring agency within AI systems. AI & society, 1-11.

[18] Shin, D., Lim, J. S., Ahmad, N., & Ibahrine, M. (2022). Understanding user sensemaking in fairness and transparency in algorithms: algorithmic sensemaking in over-the-top platform. AI & SOCIETY, 1-14.

[19] Novelli, C., Taddeo, M., & Floridi, L. (2023). Accountability in artificial intelligence: what it is and how it works. AI & SOCIETY, 1-12.

[20] Qadir, J., Islam, M. Q., & Al-Fuqaha, A. (2022). Toward accountable human-centered AI: rationale and promising directions. Journal of Information, Communication and Ethics in Society.

[21] Mohamed, M. (2023) "Agricultural Sustainability in the Age of Deep Learning: Current Trends, Challenges, and Future Trajectories", Sustainable Machine Intelligence Journal, 4, p. (2):1–20. doi: 10.61185/SMIJ.2023.44102.

[22] Muthuswamy, M. and M. Ali, A. (2023) "Sustainable Supply Chain Management in the Age of Machine Intelligence: Addressing Challenges, Capitalizing on Opportunities, and Shaping the Future Landscape", Sustainable Machine Intelligence Journal, 3. doi: 10.61185/SMIJ.2023.33103.

Received: 01 Feb 2024, **Revised:** 17 May 2024,
Accepted: 18 Jun 2024, **Available online:** 20 Jun 2024.