THE IMPACT OF LARGE LANGUAGE MODELS (LLM) ON SCIENTIFIC WRITING WITH SPECIAL REFERENCE TO CHATGPT

Dr. Pooja P. Dadhe
Asst. Librarian,
R.T.M. University, Nagpur
Email : **ppdadhe@gmail.com**

Dr. Gulshan D. Kuthe
Librarian,
Lokmanya Tilak Mahavidyalaya, Wani
e-mail:gulshankuthe30@gmail.com

---

Abstract:

Artificial Intelligence (AI) augmented Large Language Models (LLM) have made a significant impact on academic writing due to their potential to improve teaching and learning. The use of generative AI like ChatGPT has rapidly transformed the research and publication landscape, but their use in scientific writing has generated some concern predominantly on ethical aspects. Considering the mixed opinion of the scholarly world on the ethical implication of these technologies in scientific writing, making the prospective users informed is very essential. This paper aims to explore the opportunities and concerns that researchers have expressed about ChatGPT.

Key Words: Artificial Intelligence (AI), Large Language Models (LLM), Chatbots, ChatGPT, Scientific Writing

----

Introduction:

Over the past decade, advances in information and communication technologies especially artificial intelligence (AI) have transformed educational practices significantly. As a transformative technology, AI has been widely adopted across various fields of which academic writing has seen a remarkable advancement (Golan et al., 2023). Especially the use of Large Language Models (LLM) in education has been considered as a potential area of interest as it provides a diverse range of applications to make teaching and learning more effective (Kasneci et al., 2023). One of the noteworthy development of Large Language Models (LLM) is Generative Pre-Trained Transformer (GPT), which is built on the Transformer architecture. GPT is pre-trained on vast amounts of text data and has the potential of generating nearly human-like text. This technology has evolved from GPT-1 to GPT-3, and ultimately ChatGPT (Nakaura&Naganawa, 2023), (Natalie, 2023).

From the very beginning of the launch, ChatGPT has garnered worldwide attention. It was quickly acknowledged as a disruptive technology that can revolutionize the information-seeking and content-writing behavior of people across all levels. ChatGPT holds extensive capacity as a valuable and robust tool for various tasks, including automated draft creation, article summarization, language translation (Salvagno et al., 2023), (Nakaura&Naganawa, 2023) andeven computer coding (OpenAI, 2022). These applications of ChatGPT conceivably be beneficial in academic work particularly by simplifying and accelerating the writing process. ChatGPT's capabilities can help reduce the time and effort required for these tasks, making them more accessible and efficient for researchers and students alike (Salvagno et al., 2023). New AI language models are garnering a lot of interest, but what one can
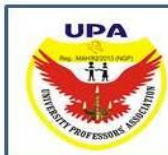
actually do with them is need to be addressed as, it is also believed that utilizing AI for specialized research is susceptible to introducing inaccuracies, bias, and plagiarism (van Dis et al., 2023). Hence, there is an urgent need for students, teachers, researchers, publishers, and especially library professionals to understand the challenges and opportunities of ChatGPT.

1. Overview of ChatGPT:

ChatGPT is a Large Language Model (LLM) developed by OpenAI which interacts with the users in a conversational way. ChatGPT (Chatbot Generative Pre-Trained Transformer) is a type of Artificial Intelligence (AI) software designed to simulate conversations with human users (Salvagno et al., 2023). The model was trained on a corpus of data available on the internet (mainly open access). This sensational language model functions as a Chatbot interface and exhibits remarkable competencies in generating human-like text in response to prompts. It was optimized for dialogue and hencehas the ability to create engaging dialogue by linking the previous conversation history with the user. It is also trained to provide follow-up corrections and even decline inappropriate prompts (Natalie, 2023). It was also trained using Reinforcement Learning with Human Feedback (RLHF), it is a technique that enables the Chatbot to understand what humans expect when they ask a question (OpenAI, 2022).Since its release in November 2022, it has taken the whole world by storm because of its disruptive potential for all most all fields starting from business firms to healthcare where academia needs special mention. It would be presumptuous to speculate what impact this technology will have on librarianship. However, its impact on the generation and publication of scientific literature is apparently unavoidable and hence needs to be addressed immediately.

2. Impact of Large Language Models (LLM) on scholarly publications:

Large Language Models (LLM) which are based on conversational Artificial Intelligence technology, have wide-ranging consequences on scholarly writing. These language models have the potential to offer customized and personalized assistance to educators, aiding them in providing adaptive learning approaches for students. This can make learning more engaging and effective (Kasneci et al., 2023). Conversational AI will conceivably transform research practices and publishing, creating both opportunities and concerns (van Dis et al., 2023). Given the apparent revolutionary impact of ChatGPT on scholarly publications, it is important to examine its efficacy and dependability of it in order to enable the researchers to use the technology judiciously for their research needs.

2.1 Opportunities:
2.1.1. Discovery search engine:

The natural language processing (NLP) capabilities of ChatGPT have changed the way people interact with search engines. It can work as an alternative to search engines like Google (Montti, 2022) and provide opportunities to find out the required information of all sorts through a single search window. ChatGPT's expertise lies in its ability to complete tasks quickly and provide the user with the personalized information which he/she looks for without visiting several web pages. And thus saves a great deal of time and effort by refraining from the huge amount of non-relevant data. With proper knowledge, researchers can use ChatGPT to enhance their productivity.

UPA NATIONAL E-JOURNAL
Interdisciplinary Peer-Reviewed Indexed Journal
Special Issue: Volume -10 : Issue-1 (January-2024)

ISSN
2455-4375

2.1.2 Language Assistance:

ChatGPT can be used as a text processing and editing tool and enhance writing skills. Studies suggest that ChatGPT can write impressive essays on diverse topics, summarize (Kasneci et al., 2023) and paraphrase research papers, translate competitively with popular translation tools (e.g., GoogleTranslate) (Jiao, Wang, Huang, & Wang, 2023). it can be used to a great extent for language learning (Lang, 2022).

2.1.3 Acceleration of scientific publication:

Generative AI could help a wide range of people mainly from non-English speaking countries to improve their language skills and increase their productivity and, consequently decreasing the performance gap between researchers. It has also the potential to influence the review and speed up the publication process (Checco et al., 2021). By helping people to write fluently and by reducing the publication time ChatGPT can accelerate the innovation process making scientific production more fair and inclusive (van Dis et al., 2023).

2.2 Challenges:

2.2.1 Inaccurate information:

The basic structure of ChatGPT, which is trained on vast amounts of data from the internet, endangers the output to be factual and accurate. The corpus of data is largely built on open-access materials, some of which may not have undergone thorough fact-checking for accuracy. Consequently, there is a substantial likelihood that LLMs may learn scientifically incorrect information, meaning there is no guarantee that their output will be accurate (Nakaura&Naganawa, 2023). Studies also suggest that Large Language Models are very convincing in producing text, but are found to be often wrong, so it is vulnerable to the spreading of misinformation, and misleading scientific facts (van Dis et al., 2023).

2.2.2 Biases:

The limitations of large language models with regard to factual information are not properly addressed yet by the research community. Hence, the probability of biases owing to prejudice of the creator of the data cannot be ruled out. Further, a self-proclaimed statement by Open AI also states that outputs provided by ChatGPT may be inaccurate, untruthful, biased, and otherwise misleading at times (Natalie, 2023).

2.2.3 Limited data:

At present ChatGPT is trained on a corpus of data covering up to 2021. And as it is not connected to the internet yet, it still functions on the corpus of the training data leading to its limited knowledge of the world and events after 2021(Joshua, 2023). Hence, the result provided by the model lacks comprehensiveness.

2.2.4 Degradation of quality:

The basic qualities of scientific writing rely heavily on novelty, creativity, transparency, fact-checking, and dependability. As contents generated by ChatGPT does not fulfill the basic criteria, it could degrade the quality of research
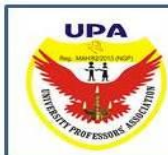
2.2.5 Copyright infringement:

The ability of ChatGPT reproduces text based on the corpus of training data brings up concerns about potential copyright violations as it lacks the citation of the original source or creator. As per copyright law, the first owner of a work's copyright is its creator. However, it is not practicable at least at this stage to determine who owns the copyright in the content generated by ChatGPT (McKendrick, 2022), (Commission, 2023). The model is trained on a vast amount of text from various sources, which could suggest that the creators of the source material should have some claim to the copyright in the generated content. Even, it could be argued that OpenAI, as the creators of the program and the ones who wrote the code that generates the content, has a great chance to hold the copyright in the future when everything will be in place. Hence, researchers using ChatGPT for scientific writing are at risk of not giving credit to earlier work, and hence unknowingly risking themselves to copyright infringement.

2.2.6 Accountability

Fixing accountability for content generated by ChatGPT is the most debatable topic in the scholarly world of late. In some instances, ChatGPT has been credited as the author of the researcher paper (King & ChatGPT, 2023), (Curtis & ChatGPT, 2023), (O'Connor & ChatGPT, 2023), raising doubt regarding the eligibility to meet authorship requirements. The argument against such authorship is based on the inability of LLMs to be held accountable as authors (Stokel-Walker, 2023).

2.2.7 Plagiarism

Tools are already in place to verify the creator of a text whether machines or humans. Though at present with advanced AI technologies and smart engagement with machines, the detection may be evaded, however, people have started working on technology that can identify text generated by ChatGPT. GPTZero is one such tool that is underway (gptzero.me/faq, n.d.)and may be like plagiarism detection tools; such tools will become standard in publishing and education (Mitra, 2023), (Barnett, 2023), (Hern, 2022), (Sharma, 2023).

3. Conclusion:

Given the potential of ChatGPT in content writing it is inevitable that the researchers will be engaged with the technology to enhance productivity. Understanding that researchers use LLMs in their work, librarians, teachers, editors, and publishers need to remain cautious. Librarians along with teachers can make the researchers aware of how to use the technology judiciously for their research needs in order to reduce their biases and improve productivity. LLMs are evolving at a very fast rate and hence there is an urgent need for guidelines from competent authorities, that can raise awareness of the use of such technology in education and research. Library and Information Science professionals should welcome the AI revolution with open arms and explore the opportunities in developing technology-driven services to support their users.

References:

- Barnett, S. (2023, January). *ChatGPT Is Making Universities Rethink Plagiarism*. Retrieved March 25, 2023, from www.wired.com: https://www.wired.com/story/chatgpt-college-university-plagiarism/

- Checco, A., Bracciale, L., Loreti, P., Pinfield, S., & Bianchi, G. (2021). AI-assisted peer review. *Humanities and Social Sciences Communications*, *8*(1), Article 1. https://doi.org/10.1057/s41599-020-00703-8

- Commission, E. (2023, February 20). *Intellectual Property in ChatGPT*. Retrieved March 21, 2023, from intellectual-property-helpdesk.ec.europa.eu: https://intellectual-property-helpdesk.ec.europa.eu/news-events/news/intellectual-property-chatgpt-2023-02-20_en

- Curtis, N., & ChatGPT. (2023). To ChatGPT or not to ChatGPT? The Impact of Artificial Intelligence on Academic Publishing. *Pediatric Infectious Disease Journal*, *42*(4), 275–275. https://doi.org/10.1097/INF.0000000000003852

- Golan, R., Reddy, R., Muthigi, A., & Ramasamy, R. (2023). Artificial intelligence in academic writing: A paradigm-shifting technological advance. *Nature Reviews Urology*, 1–2. https://doi.org/10.1038/s41585-023-00746-x

- *gptzero.me/faq*. (n.d.). Retrieved March 22, 2023, from https://gptzero.me/faq: https://gptzero.me/faq

- Hern, A. (2022, December 31). *AI-assisted plagiarism? ChatGPT bot says it has an answer for that*. Retrieved Month 25, 2023, from www.theguardian.com: https://www.theguardian.com/technology/2022/dec/31/ai-assisted-plagiarism-chatgpt-bot-says-it-has-an-answer-for-that

- Jiao, W., Wang, W., Huang, J.-t., & Wang, v. X. (2023, March 19). *Is ChatGPT A Good Translator? Yes With GPT-4 As The Engine*. Retrieved March 22, 2023, from arxiv.org: https://arxiv.org/pdf/2301.08745.pdf

- Joshua, J. (2023). *Why doesn't ChatGPT know about X?* Retrieved March 28, 2023, from help.openai.com: https://help.openai.com/en/articles/6827058-why-doesn-t-chatgpt-know-about-x

- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

- King, M. R. & chatGPT. (2023). A Conversation on Artificial Intelligence, Chatbots, and Plagiarism in Higher Education. *Cellular and Molecular Bioengineering*, *16*(1), 1–2. https://doi.org/10.1007/s12195-022-00754-8

- Lang, F. (2022, December 15). Using ChatGPT as your personal language learning assistant. *Medium*. https://medium.com/@langfab/using-chatgpt-as-your-personal-language-learning-assistant-bdb03bef8565

- McKendrick, J. (2022, December 21). *Who Ultimately Owns Content Generated By ChatGPT And Other AI Platforms?* Retrieved March 22, 2023, from www.forbes.com: https://www.forbes.com/sites/joemckendrick/2022/12/21/who-ultimately-owns-content-generated-by-chatgpt-and-other-ai-platforms/?sh=fde9e1c5423a