Artificial Intelligence in Scientific Writing: A Deuteragonistic Role?

Benny Wohlfarth [1] , Samuel R. Streit [2] , Sissel Guttormsen [3]

1. Institute for Medical Education, University of Bern, Bern, CHE 2. Department of Angiology, Inselspital Bern, Bern University Hospital, Bern, CHE 3. Institute of Medical Education, University of Bern, Bern, CHE

Corresponding author: Benny Wohlfarth, benny.wohlfarth@protonmail.com

Abstract

In this article, we reflect on the pros and cons of artificial intelligence (AI)-augmented scientific writing for more comprehensible research towards society to gain trust for science-led policy. For this purpose, we integrated our thoughts into the *Factors of Perceived Trustworthiness from Mayer, Davis, and Schoorman's Model of Trust* and made propositions to define AI's role in trustful scholarly communication.

Categories: Other
Keywords: ai regulation, large language models (llms), chatbots, swa (scientific writing assistance), research integrity

Editorial

Letter to the editor

In ancient Greek drama, the deuteragonist is the second actor in the play. While interacting closely with the protagonist to support and assist, the deuteragonist can also assume a role of conflict or rivalry; ultimately helping to develop the main character and shape the overarching narrative of the play [1].

Since artificial intelligence (AI) came to the stage in 1956 [2], it has gained momentum from the 1990s onwards. Today, it affects nearly every aspect of how we live and work as a society, whether in healthcare [3], education [4], economics [5], arts [6], or science [7]. Generally, people have high trust in science [8]. However, this trust is being tested due to incidents of human bias, error, and fraud in production [9-11], interpretation [11,12], and reporting and publication of scientific work [12,13]. Particularly since the COVID-19 pandemic, conflicting scientific results have done little to cultivate greater trust in science [14,15]. For society's acceptance and trust in science to grow, it is essential to make research as transparent and comprehensible as possible to everyone in the "theatron," while maintaining robustness and objectivity [16].

Balancing the benefits and risks of AI in scholarly communication

The concept of AI assisting researchers in scientific writing is appealing. If executed correctly, AI could support more comprehensible research and increase acceptance of science-led policy by the general public. However, concerns have been raised about issues such as plagiarism and obscurity in research, particularly with unregulated use of AI-driven scientific writing assistance (SWA) and chatbots [17,18]. In this article, we reflect on the pros and cons of AI in scientific writing. To do so, we integrate these considerations into the Factors of Perceived Trustworthiness from Mayer, Davis, and Schoorman's Model of Trust (see Figure *1*) [19]. Originating from the social sciences, this model has proven valuable and robust across various disciplines and is also widely utilized in information systems-related science [20,21].
This background renders it perfect to transfer our thoughts on AI's role in scientific writing.

FIGURE 1: Factors of Perceived Trustworthiness from Mayer, Davis, and Schoorman's Model of Trust

Ability refers to skills, competencies, and characteristics that enable an individual to be influential in a particular domain. *Benevolence* holds the idea that the trustee has a genuine intention to do good for the trustor, independent of self-centered motives. *Integrity* is the trustor's perception that the trustee adheres to a set of principles deemed acceptable by the trustor [19].

The pros

SWA has the potential to make research more accessible and understandable by facilitating core findings for wider target groups, including non-scientific stakeholders, policymakers, and the public (benevolence+) [22]. In addition, SWA and chatbots hold a chance to offer more complex language assistance to non-native English speakers, who may struggle with the costs of language editing services (benevolence+) [23]. This augmentation further improves scientific writing by suggesting simple yet correct sentence structures, enhancing clarity and the overall readability of a manuscript (ability+).

AI can generate various visual representations of research, ranging from simple semiotic visualizations to more complex schematic views. These offer effective and easily comprehensible data communication for other researchers and a broader audience in a non-textual form (ability+).

Humans are prone to errors, especially when performing repetitive and monotonous tasks like formatting citations and proofreading [24]. AI automation for such tasks counteracts decision fatigue and frees time for scientists to focus on the research content, specifically, interpreting and concluding from data, thereby ultimately accelerating scientific progress for society (ability+) [25].

AI can quickly analyze and summarize vast amounts of scientific literature or content data, enabling researchers to uncover patterns, correlations, and trends not easily identifiable through manual analysis alone (ability+). This can also contribute to improve the overall quality of the research process; for example, in identifying relevant research questions and interpreting data, resulting in a more targeted design of original research and literature reviews (benevolence+) [26,27].

The cons and possible measures

Since AI chatbots are based on large language models (LLM), their answers are always a function or a derivate of their repertoire and hold a risk of bias to scientific writing, based on possible biases in the training datasets themselves. This may include racial, sexual, or religious bias, making the results less inclusive [28-30] (integrity⁻). This should be addressed by fostering the establishment of balanced training of datasets, bias detection, and correction algorithms, as well as close human supervision and feedback mechanisms [29,31,32].

Automation of tasks, e.g., the aforementioned SWA, may stoke fear of job loss [33] (benevolence⁻). Mitigating such concerns must be enforced by public dialogue, addressing integrative scenarios including supervising, and collaborating definitions as well as adequate education and reskilling programs for the people in an increasingly AI-automated world.

The crystalline nature of LLMs precludes chatbots from actuality. More importantly, while they are neither living nor experiencing culture, LLM chatbots are put into an insurmountable epistemic distance toward the user's world of thought, which ultimately makes current chatbots incapable of generating a priori ideas or conclusions for society (integrity⁻, ability⁻). This could be closed by the aforementioned human-AI collaboration, as well as pairing LLMs with augmented reality, evolutionary algorithms, generative models, fine-tuning, and zero-/few-shot learning to generate inclusive novel ideas and writings [34-36].

Overreliance on AI for writing tasks can lead to a decline in critical thinking and writing skills among researchers, as well as an impoverishment of creativity and inclusivity. This could potentially jeopardize the integrity of scientific research (integrity⁻), lowering people's trust in it, and open the door for predatory journals to misuse AI to mass-produce low-quality, inaccurate, and misleading content, based on monetary interests [37] (benevolence⁻). We suggest the encouragement of education and training in young researchers on critical thinking and writing skills, highlighting that AI is a tool of assistance, not a replacement for the human intellect [37,38] and applaud the viewpoint to see AI more in the sense of an "augmented intelligence" rather than an "artificial intelligence" by successfully partnering with machine intelligence [39]. Furthermore, rules and standards for the use of AI in scientific writing must be agreed on to maintain research integrity and discourage misuse, with possible penalties for infractions [37,38]. The development of advanced AI auditing systems to detect and prevent the mass production of low-quality, inaccurate, or misleading content is needed. Furthermore, strict peer review processes must be a standard, as well as automated checks against known databases. The latter is already possible through various tools that require mandatory human supervision [23,40].

Lack of explainability [41] and the generation of misleading or hypothetical claims and references, known as hallucination [42] are critical topics in regard to society's trust. Unverified adoption of such texts and references can result in legal consequences and ultimately produce false results and conclusions that impede society's trust in research (integrity⁻). We advise fact-checking and the use of verification systems to confirm the accuracy of AI-generated text before publication or use, predominantly by human reviewers, but also other AI-based systems, as well as the development of AI algorithms specifically designed to detect and correct hallucinations and other errors in AI-generated text [43].

Implementation and effective utilization of AI in scientific writing requires technical expertise and resources. Not all researchers may have access to or be comfortable with using it, potentially creating a split in the scientific community (integrity⁻). Additionally, access to AI tools may be limited due to factors such as cost or institutional constraints (ability⁻, integrity⁻). This could be counteracted by faculties by providing comprehensive training to researchers on the use of AI tools, ensuring that they are comfortable and proficient in their application.

In general, governmental and international organizations should foster the development and promotion of open-source AI models like BLOOM, which holds the opportunity to make AI-based research more inclusive by reducing potential splits caused by financial or institutional constraints.

Future plays: regulation, not restriction

Famous misjudges in the past show that new technology is often exposed to cognitive bias, such as status quo bias [44]. Even experts and opinion leaders plunged into false predictions and disputed the advantages of new technology throughout history [45]. In this sense, the outright rejection of AI in scientific writing is not a solution. Undoubtedly, AI is here and is gaining terrain quickly. OpenAI's ChatGPT, which was launched in November 2022, exemplifies this technology and its rapid development. With more than 100 million users in the first two months of its release, it is the fastest-growing application in the history of web applications [46].

Ethical Considerations

To define AI's role in the future of scientific writing, our reflection on the pros and cons, combined with suggestions for reducing risks, is an important contribution to this debate. Considering that AI presumably will become more complex and involved in the future, e.g., in the concept of artificial general intelligence (AGI) [34,35], it is crucial to discuss and find consensus on the raised concerns in a timely manner. We see the potential for AI in scientific writing, given that robust and ethical frameworks for the incorporation into the daily work of scientists can be established. After all, it must preserve the commitment for scientific integrity and trustworthiness towards society and at the same time acknowledge the thirst for innovation.

Limitations of AI for Scientific Writing

The inherent risk of introducing bias, the phenomenon of "hallucination," a lack of originality, limited accessibility, and insufficient transparency and explainability hinder AI from becoming a protagonist in scientific writing at its current technological stage and without clear guidelines.

For now, we advocate the general disclosure of AI usage in the methods section of research articles as an initial step. The declaration should include the type of AI used and clarify the role the AI had in the creation of the manuscript. In addition to that, the corresponding scripts and prompts, as well as chatbot parameters, should accompany the article as supplementals and data repositories. To guarantee full transparency, the following measures should be developed and established: (a) standardized reporting frameworks customized to the AI model in use. This ensures that essential AI-related details are reported consistently and thoroughly, providing additional clarity to the reader. (b) Introduce a visual cue on the published article, in analogy to open access labeling, by creating an international AI usage badge (Figure 2); a visual indicator displayed on studies employing AI, signifying adherence to AI good practices. Akin to the open-access badges, this would allow immediate recognition of AI's use and its compliance with standard transparency guidelines. (c) Scientific databases should have a search option to include or exclude papers with AI involvement in the making of the manuscript. (d) We suggest the introduction of an "AI Statement" section. This would include a comprehensive report detailing why AI was used, its assigned tasks, possible biases of the model, mitigation strategies, and how ethical considerations were addressed. In the sense of multidisciplinarity and in analogy to statisticians, being involved in large and sophisticated analyses, this may include an AI responsible person with an appropriate technical background when AI is heavily involved.
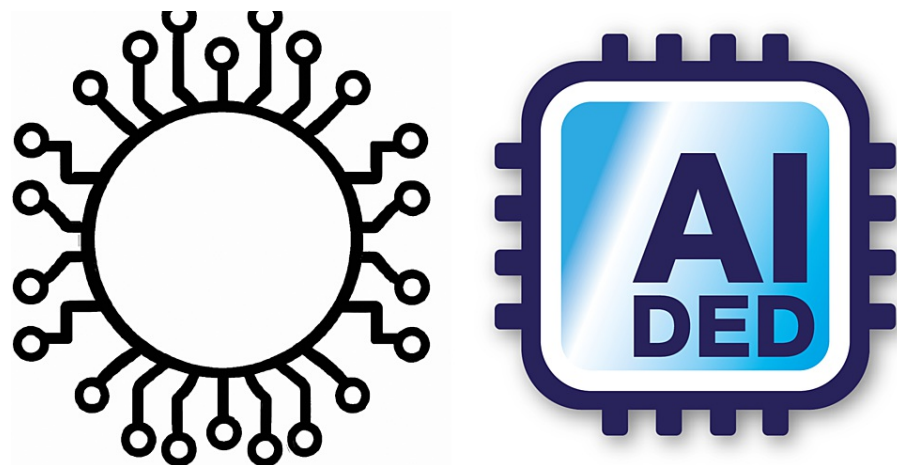

FIGURE 2: Proposition of badges for labeling AI usage in scientific writing.

Left: AI-generated badge by DALL·E on the prompt: "Please make a symbolic logo that could be used as a badge to label text passages in scientific text, that have been generated by AI", not quite meeting cultural expectations, as the circuits could be negatively associated, for example, with the resemblance of Medusa's hair. Right: AIDED badge, human creation done in Adobe Illustrator (Adobe Inc., San Jose, California, United States) by the authors.

Conclusions

The boundary that separates AI as a mere tool and a potential autonomous contributor - a protagonist, akin to a coauthor - gradually becomes less distinct as its technology and usage evolve. However, the future role of AI as a protagonist in research has to be postponed. For now, AI will remain in the second line, making us protagonists better authors through its help and our consecutive reflection on the rules of good scientific writing for a more comprehensible and trustful scholarly communication.

Additional Information

Disclosures

References

1. Easterling P: The Cambridge Companion to Greek Tragedy . Cambridge University Press, Cambridge; 1997. 10.1017/CCOL0521412455
2. McCarthy J, Minsky M, Rochester N, Shannon C: A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. Dartmouth College, Hanover; 1955. 10.1609/aimag.v27i4.1904
3. Rajpurkar P, Chen E, Banerjee O, Topol EJ: AI in health and medicine . Nat Med. 2022, 28:31-8. 10.1038/s41591-021-01614-0
4. Chiu TKF, Xia Q, Zhou X, Chai CS, Cheng M: Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. CAEAI. 2023, 4:100118. 10.1016/j.caeai.2022.100118
5. Loureiro SMC, Guerreiro J, Tussyadiah I: Artificial intelligence in business: state of the art and future research agenda. J Bus Res. 2021, 129:911-26. 10.1016/j.jbusres.2020.11.001
6. Sony World Photography Awards 2023. (2023). Accessed: July 16, 2023: https://www.artforum.com/news/sony-world-photography-award-winner-reveals-entry-was-ai-generated-rejects-prize-90416.
7. Xu Y, Liu X, Cao X, et al.: Artificial intelligence: a powerful paradigm for scientific research . Innovation (Camb). 2021, 2:100179. 10.1016/j.xinn.2021.100179
8. Hendriks F, Kienhues D, Bromme R: Trust in science and the science of trust . Trust and Communication in a Digitized World. Progress in IS. Blöbaum B (ed): Springer, Cham; 2016. 143-59. 10.1007/978-3-319-28059-2_8
9. George SL: Research misconduct and data fraud in clinical trials: prevalence and causal factors . Int J Clin Oncol. 2016, 21:15-21. 10.1007/s10147-015-0887-3
10. George SL, Buyse M: Data fraud in clinical trials . Clin Investig (Lond). 2015, 5:161-73. 10.4155/cli.14.116
11. Andrade C: HARKing, cherry-picking, P-hacking, fishing expeditions, and data dredging and mining as questionable research practices. J Clin Psychiatry. 2021, 82:20f13804. 10.4088/JCP.20f13804
12. Wicherts JM: The weak spots in contemporary science (and how to fix them) . Animals (Basel). 2017, 7:90. 10.3390/ani7120090
13. Song F, Parekh S, Hooper L, et al.: Dissemination and publication of research findings: an updated review of related biases. Health Technol Assess. 2010, 14:iii, ix-xi, 1-193. 10.3310/hta14080
14. Abbasi J: The flawed science of antibody testing for SARS-CoV-2 immunity . JAMA. 2021, 326:1781-2. 10.1001/jama.2021.18919
15. Alves CP, Barreto Segundo JD, da Costa GG, Pereira-Cenci T, Lima KC, Demarco FF, Crochemore-Silva I: How a few poorly designed COVID-19 studies may have contributed to misinformation in Brazil: the case for evidence-based communication of science. BMJ Open Sci. 2021, 5:e100202. 10.1136/bmjos-2021-100202
16. Bennett M: Should I do as I'm told? Trust, experts, and COVID-19 . Kennedy Inst Ethics J. 2020, 25:243-63. 10.1353/ken.2020.0014
17. Thorp HH: ChatGPT is fun, but not an author . Science. 2023, 379:313. 10.1126/science.adg7879
18. Flanagin A, Bibbins-Domingo K, Berkwits M, Christiansen SL: Nonhuman "authors" and implications for the integrity of scientific publication and medical knowledge. JAMA. 2023, 329:637-9. 10.1001/jama.2023.1344
19. Mayer RC, Davis JH, Schoorman FD: An integrative model of organizational trust. Acad Manage Rev. 1995, 20:709. 10.5465/amr.1995.9508080335
20. Grimm R, Maier M, Rothmund T: Vertrauen. Datenschutz und Datensicherheit. 2015, 5:283-8. 10.1007/s11623-015-0414-8
21. Söllner M, Hoffmann A, Hoffmann H, Wacker A, Leimeister JM: Understanding the formation of trust in IT artifacts. Proceedings of the International Conference on Information Systems (ICIS). Orlando, Florida; 2012. 10.1007/978-3-319-05044-7_3
22. Sinclair BJ: Letting ChatGPT do your science is fraudulent (and a bad idea), but AI-generated text can enhance inclusiveness in publishing. Curr Res Insect Sci. 2023, 3:100057. 10.1016/j.cris.2023.100057
23. Nazari N, Shabbir MS, Setiawan R: Application of artificial intelligence powered digital writing assistant in higher education: randomized controlled trial. Heliyon. 2021, 7:e07014. 10.1016/j.heliyon.2021.e07014
24. Langner R, Eickhoff SB: Sustaining attention to simple tasks: a meta-analytic review of the neural mechanisms of vigilant attention. Psychol Bull. 2013, 139:870-900. 10.1037/a0030694
25. Pignatiello GA, Martin RJ, Hickman RL Jr: Decision fatigue: A conceptual analysis . J Health Psychol. 2020, 25:123-35. 10.1177/1359105318763510
26. Huang J, Tan M: The role of ChatGPT in scientific communication: writing better scientific review articles . Am J Cancer Res. 2023, 13:1148-54.
27. Fatani B: ChatGPT for future medical and dental research . Cureus. 2023, 15:e37285. 10.7759/cureus.37285
28. Dhamala J, Sun T, Kumar V, Krishna S, Pruksachatkun Y, Chang KW, Gupta R: BOLD: dataset and metrics for measuring biases in open-ended language generation. FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. Association for Computing Machinery, New York, NY; 2021. 862-72. 10.1145/3442188.3445924
29. Liang PP, Wu C, Morency LP, Salakhutdinov R: Towards understanding and mitigating social biases in language models. Proceedings of the 38th International Conference on Machine Learning. Proceedings of Machine Learning Research, 2021. 6565-76. 10.48550/arXiv.2106.13219

2023 Wohlfarth et al. Cureus 15(9): e45513. DOI 10.7759/cureus.45513

5 of 6

30. Belenguer L: AI bias: exploring discriminatory algorithmic decision-making models and the application of possible machine-centric solutions adapted from the pharmaceutical industry. AI Ethics. 2022, 2:771-87. 10.1007/s43681-022-00138-8

31. Ouyang L, Wu J, Jiang X, et al.: Training language models to follow instructions with human feedback. ArXiv. 2022, 10.48550/arXiv.2203.02155

32. European Union Agency for Fundamental Rights: Bias in algorithms - artificial intelligence and discrimination. European Union Agency for Fundamental Rights, Vienna, Austria; 2022. https://fra.europa.eu/en/publication/2022/bias-algorithm.

33. Deranty JP, Corbin T: Artificial intelligence and work: a critical review of recent research from the social sciences. AI Soc. 2022, 10.1007/s00146-022-01496-x

34. Fei N, Lu Z, Gao Y, et al.: Towards artificial general intelligence via a multimodal foundation model. Nat Commun. 2022, 13:3094. 10.1038/s41467-022-30761-2

35. Chen Y, Qian X, Ceng T: A preliminary study on the capability boundary of LLM and a new path towards AGI. 2023.

36. Banerjee D, Poser M, Wiethof C, et al.: A system for human-AI collaboration for online customer support. The AAAI 2023 Workshop on Representation Learning for Responsible Human-Centric AI (R2HCAI). Association for the Advancement of Artificial Intelligence (AAAI), Washington, DC; 2023.

37. van Dis EA, Bollen J, Zuidema W, van Rooij R, Bockting CL: ChatGPT: five priorities for research. Nature. 2023, 614:224-6. 10.1038/d41586-023-00288-7

38. Dergaa I, Chamari K, Zmijewski P, Ben Saad H: From human writing to artificial intelligence generated text: examining the prospects and potential threats of ChatGPT in academic writing. Biol Sport. 2023, 40:615-22. 10.5114/biolsport.2023.125623

39. Lomis K, Whelan A, Sheikh J, James C, Chan TM: Artificial intelligence: what medical educators should be doing now. AMEE Lyon 2022. Association for Medical Education in Europe (AMEE), Dundee, Scotland; 2022.

40. Desaire H, Chua AE, Isom M, Jarosova R, Hua D: ChatGPT or academic scientist? Distinguishing authorship with over 99% accuracy using off-the-shelf machine learning tools. Cell Rep Phys Sci. 2023, 30:4. 10.48550/arXiv.2303.16352

41. Amann J, Blasimme A, Vayena E, Frey D, Madai VI: Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med Inform Decis Mak. 2020, 20:310. 10.1186/s12911-020-01332-6

42. Athaluri SA, Manthena SV, Kesapragada VS, Yarlagadda V, Dave T, Duddumpudi RT: Exploring the boundaries of reality: investigating the phenomenon of artificial intelligence hallucination in scientific writing through ChatGPT references. Cureus. 2023, 15:e37432. 10.7759/cureus.37432

43. Birhane A, Kasirzadeh A, Leslie D, Wachter S: Science in the age of large language models. Nat Rev Phys. 2023, 5:277-80. 10.1038/s42254-023-00581-4

44. Godefroid ME, Plattfaut R, Niehaves B: How to measure the status quo bias? A review of current literature. Manag Rev Q. 2022, 28:1-45. 10.1007/s11301-022-00283-8

45. Naaz S, Asghar A: Artificial intelligence, nano-technology and genomic medicine: the future of anaesthesia. J Anaesthesiol Clin Pharmacol. 2022, 38:11-7. 10.4103/joacp.JOACP_139_20

46. ChatGPT is the fastest growing app in the history of web applications. (2023). Accessed: June 28, 2023: https://www.forbes.com/sites/cindygordon/2023/02/02/chatgpt-is-the-fastest-growing-ap-in-the-history-of-web-applicati....