Responsible but Innovative Use of Artificial Intelligence in Scientific Publishing

Trevor Champagne[1,2] and Neil Shear[2]

In late November 2022, what is arguably the first widely accessible large language model (LLM), ChatGPT (chat.openai.com, 2023), was released for public use and appraisal. This represents a step on the road toward artificial intelligence (AI) by statistically mimicking human-generated factual and stylistic content using large mathematical models that have been trained on websites, in books, and in other media and furthermore curated and corrected by humans working with the OpenAI team. There have been many enthusiastically proposed use cases in the medical field and the release of preliminary pilot products, such as AI medical scribes, and equally a number of concerns raised involving accuracy, security, privacy, and ownership.

The appeal of AI is in the near-immediate production of large amounts of plausible text, which works well for ubiquitous topics and less so for many others. Numerous articles and reports demonstrate that it is factually incorrect for many critical medical questions (Sallam, 2023). Furthermore, by design, it is unexplainable, and it cannot trace the factual source of the text it produces (Amann et al, 2020)—a gaping large problem with AI! Both issues arise owing to the fundamental nature of the underpinning transformer algorithms: these models are built not through knowledge but by statistically evaluating billions of samples of text to simply answer: given a series of words, what word is most likely to come next? It is clear that without additional significant human intervention, ChatGPT (and other LLMs) is not a source, and you cannot use ChatGPT to suggest sources—there is a real risk that they will be imaginary.

This risk of imaginary data (also called hallucinations, derived from the temperature of the algorithm) is inseparable from what makes the system attractive in the first place. Asking it to summarize an article, to elaborate on a sentence, or to write a story requires a degree of freedom to create text beyond rote plagiarism. In this sense, philosophically, it is acting as a creative agent through the mechanism of statistics.

We need to know more about what AI can do—it is possible that the statistical models when pressed can produce new considerations regarding known diseases or novel ways to diagnose or treat conditions or act as a hypothesis generator in research. It is true that forming new connections between previously unrelated pieces of data is a fundamental foundation of human creativity!

Small groups seeking publication of novel research face many barriers and pitfalls. Submitting to multiple journals with different specific word counts can be one laborious task that LLMs can do instantly. Sound research submitted to journals that are not in the author's native language can be overlooked when the article does not appropriately match the tone and timbre of the journal—another task that LLMs seem to excel at. These have the potential to unlock and promote new research globally for researchers who could not otherwise afford a team of medical writers and editors. The expectation should be similar to the rigor of clinical trials—that the use of an LLM should be disclosed and the inputs, the queries, and the outputs should be all readily available as appendices for scrutiny or further analysis.

Encouraging this form of open communication will further the truthful, accepted use of LLMs, but there are worrisome considerations. Over time, new LLMs being trained on new data and information drawn from the Internet run the risk of being trained on LLM-generated material in a vicious feedback cycle of hallucinatory nonsense (Shumailov et al, 2023[1]). Stated more explicitly, when an LLM is training for what word should come next in this sequence, if it is training from text that was already generated by another imperfect LLM generating imperfect sequences, these studies suggest that it leads inexorably into absurdity. If we instead want LLMs to improve, especially on a factual basis, the trainers need to know what sources are reliable human-generated content and what are not. It is not clear whether we will be able to rely on the tools to do it: currently, GPT-4 is not necessarily able to clearly differentiate the text it has generated from human text (Bhattacharjee and Liu, 2023[2]).

One reaches further and reimagines how this can be used for guidelines or in information-producing structures such as network meta-analyses. As these models become more sophisticated and capable of better factual interpretation, we will need a formal system for determining how to safely integrate these into our discourse—should an LLM specifically trained on the acne data be a participatory member in acne consensus meetings? Should there be a Delphi-like consensus of LLMs as the AI representative in guideline

---

[1]Department of Dermatology and Skin Science, University of British Columbia, Vancouver, Canada; and [2]Division of Dermatology, Department of Medicine, University of Toronto, Toronto, Canada

Correspondence: Trevor Champagne, Department of Dermatology and Skin Science, University of British Columbia, 595 Columbia Street W, Unit 204, Kamloops V2E 1C7, Canada. E-mail: trevor.champagne@mail.utoronto.ca

[1] Shumailov I, Shumaylov Z, Zhao Y, Gal Y, Papernot N, Anderson R. Model dementia: generated data makes models forget. arXiv 2023.
[2] Bhattacharjee A, Liu H. Fighting fire with fire: can ChatGPT detect AI-generated text? arXiv 2023.

COMMENTARY

creation? This will be an interesting and challenging direction for the coming decades.

When we assess the current climate surrounding AI-generated material, there are certainly many positive and negative extreme opinions. Unless we encourage transparent exploration of these new technologies, we will be driven by competing corporate influences of hype and fear until the future overtakes us—or if ChatGPT is prompted on that sentence to "please rephrase and shorten to be more neutral"—"without transparent exploration of new technologies, we may be swayed by corporate agendas as the future progresses."

ORCIDs

Trevor Champagne: http://orcid.org/0000-0002-7743-5403
Neil Shear: http://orcid.org/0000-0001-9151-1145

CONFLICT OF INTEREST

TC has 2 patents pending relating to dermatology and artificial intelligence (neither are related to large language models) and offers consulting services to industry on the feasibility of artificial intelligence initiatives in dermatology.

AUTHOR CONTRIBUTIONS

Conceptualization: TC, NS; Supervision: NS; Writing — Original Draft Preparation: TC, NS; Writing — Review and Editing: TC, NS

DISCLAIMER

Declaration of artificial intelligence and large language models. No AI or LLM tools were used to prepare this manuscript.

REFERENCES

Amann J, Blasimme A, Vayena E, Frey D, Madai VI. Precise4Q consortium. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med Inform Decis Mak 2020;20:310.

chat.openai.com. ChatGPT. https://chat.openai.com; 2023. (accessed September 1, 2023).

Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. Healthcare (Basel) 2023;11:887.