Editorial

The AI writing on the wall

🔴 Check for updates

Guidelines are urgently needed for the use of generative AI tools like ChatGPT in scientific writing.

We speculated about the start of a new chapter in artificial writing two and a half years ago[1], but as ChatGPT has been unleashed on the world there is no longer any doubt that AI generative tools are changing the landscape of scientific writing. ChatGPT is a variation of GPT-3, the 175-billion parameter large language model from OpenAI that was trained on much of the available data on the internet (up to 2021). It works as a chatbot interface and is astonishingly good at producing human-level text upon prompts, in the form of engaging dialogue, language translations, articles, poems, essays and much more, in a range of styles on demand.

Among many industries affected by ChatGPT, scientific publishing is one that needs to address potential implications soon[2]. The risk of an imminent flood of AI-generated papers where the distinction between original human-written and AI-written or rehashed content will be blurred seems real. Scientific publishers, including Springer Nature, have in recent years adopted a range of software tools to fight malpractices including paper mills, fabricated results, duplicate submissions and plagiarism. It has become an essential, largely underappreciated part of science publishing to carry out various quality checks such as whether authors and affiliations actually exist and whether parts of the text have been previously published elsewhere. ChatGPT's ability to produce large amounts of plausible-sounding content and to rewrite existing text in different styles, making plagiarism detection near-impossible, may stretch the current system to its limits and undermine trust. Efforts are underway to create apps, such as GPTZero, that can detect whether text was generated by ChatGPT. Perhaps such tools will become standard, like the plagiarism detection services often used in publishing and education. However, language models will continue to develop and existing software to detect misuse may not be able to catch up in good time.

There are certainly also promising positive applications for generative tools like ChatGPT in scientific writing and publishing. They can be used to improve the readability of advanced drafts in an editing pass, which can help non-native speakers, and could be employed in writing summaries of a scientific text. Some may find the tool useful for brainstorming, by bouncing ideas with the chatbot in a conversational way. However, these features should be used with caution and ChatGPT-edited text needs to be carefully checked as the tool cannot be trusted to get facts right or produce reliable references.

Another concern in scientific writing is that a user's prompt may generate text from ChatGPT that includes content that the user does not understand, but which the user may be tempted to incorporate into their writing. Used judiciously, this may be a productive way to learn about a topic. A downside is that ChatGPT may normalize a new form of writing in which the human user merely curates large swaths of text by rearranging the output from multiple prompts. A related issue is that ChatGPT should not be expected to produce useful new insights or offer original, stimulating views; after all, large language models produce output by combining existing data and coming up with what is statistically the average opinion. The model has no real understanding of the world, no motivations and no moral compass.

There are also concerns about GPT harbouring bias, picked up from the large amounts of data it has been trained on. Some data curation is in place and toxicity filters are added, but limitations of large language models with regard to propagating bias are not clearly understood[3]. In fact, ChatGPT is currently in a 'research preview' stage, during which the tool is free to use, so that the world is essentially taking part in a large experiment to learn about the tool's 'strengths and weaknesses'.

There is a certain element of exaggeration in the media and in online discussions regarding the disruptive potential of generative AI tools, but large language models are here to stay and will mean change in certain areas, such as in education where ChatGPT is redefining the practice of essay writing assignments. Recently, ICML, a main machine learning conference, has announced a ban on papers that contain text entirely written by a large language model, although the organizers add that the policy may evolve when the impacts of large language models on scientific publishing are better understood. New guidelines are needed in scientific publishing to resolve legitimate and unwanted applications of AI-generative tools, and we will be actively involved in discussions about this.

Published online: 20 January 2023

References
1. *Nat. Mach. Intell.* **2**, 419 (2020).
2. Grimaldi, G. & Ehrler, B. *ACS Energy Lett.* **8**, 878–880 (2023).
3. Bender, E. M., Gebru, T., McMillan-Major, A., Shmitchell, S. in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* 610–623 (2021).