https://doi.org/10.1038/s42256-023-00678-6

Writing the rules in AI-assisted writing

As many authors are experimenting with using large language models in writing articles, some guidelines are becoming clear, but these will need to evolve as the capabilities and integration of such tools develop further.

We wrote in January 2023 (ref. 1) about the possible impact of large language models (LLMs) on scientific writing and, like many others[2], we called for a community-wide discussion on guidelines for authors, publishers and others involved in the publication process to make sensible use of these new tools. Now, a few months and many media discussions on the topic later, the development and wide adoption of LLM tools continues apace. Clear guidelines for authors seem to be a moving target, but some messages have become clear.

A first point to realize is that LLM tools will soon be integrated into a wide range of standard services and applications. Microsoft has already integrated a version of GPT-4 into its Edge browser as the new Bing and plans a wider integration into Office 365. Google has similar plans for integrating generative artificial intelligence (AI) tools into various of its Workspace applications such as Google Docs and Gmail. Soon, it might be hard to avoid writing any kind of text without being offered the option to quickly and easily have the text processed by an LLM.

However, the writing or re-writing of text by LLMs can introduce incorrect but confident-sounding statements[3]. For instance, ChatGPT (the LLM-based chatbot from OpenAI) has been found to make up plausible-sounding but fake academic references. And even when actual sources are summarized, as in the new launch demo of ChatGPT-powered Bing, the content of the sources might be misrepresented. When most of the output is correct, inaccuracies can be hard to identify, and as models become more accurate, the problem could be made worse, as users may be tempted to skip fact-checking, while errors will still occur.

A clear rule for authors is that they should not blindly adopt text suggested by LLMs and need to diligently check facts and references. Moreover, as LLM tools develop and are more routinely used, authors should avoid incorporating generated text that sounds plausible without making sure they fully understand and agree with it.

While LLM tools and their integration are still evolving, it is a good practice for authors to be transparent about whether and how they have used an LLM tool in their writing. Earlier this year, the Association for Computational Linguistics (ACL) announced a policy on AI writing assistance, and stated that they introduced an additional question on their checklist for authors, to describe if such tools were used in any way. As authors are experimenting with using conversations with ChatGPT when writing articles, it can be a useful form of transparency to provide a transcript of the corresponding prompts and answers in a supplementary section, as authors of a Comment in this issue have done.

Such practices may well change as LLM tools become standard and integrated into science communication workflows, but as the advantages and downsides of their use are being explored, the science community will benefit from transparency in AI-assisted writing. In this light, once the integration of LLM tools into browsers, word processors and other applications has been finalized by Google, Microsoft and others, it would be wise for authors to opt out of default use of such language tools, so that it remains clear at which stage LLMs were used.

Another message for users is that they need to be aware that any text inserted into ChatGPT or other LLM tools is no longer private. This could become more of a concern with the availability of tools such as ChatPDF, which makes it easier for individual users to feed large parts of text into ChatGPT. By default, all ChatGPT conversations are potential training data in possession of OpenAI, and this has already led to Samsung's banning the use of ChatGPT and other generative tools, after company confidential information was leaked. Although it is possible for users to opt out of having their data collected for training, this does not oblige OpenAI to treat the input as confidential, and the data might be used in other ways in line with their privacy policy. Third parties may offer services that rely on APIs, and data provided to such a service will then be available to both the third-party company and the platform offering the LLM API (OpenAI in the case of ChatGPT). Both can use the data to improve their services or even monetize the data directly.

Finally, the question of what copyright issues apply to AI-generated content remains unclear and guidelines from policymakers are urgently needed. Although text and data mining in itself is generally not considered copyright infringement, specific output from a generative AI model may be directly related to existing work protected by copyright. However, there is no straightforward way to identify such cases. A related open question is who owns the copyright of AI-generated work. In a recent development, the European AI Act, which has been under construction for some years, has been updated with rules for generative AI, including the requirement to disclose the use of copyrighted material in the training data. How Google, Microsoft and other companies racing to integrate LLMs in their products will adapt to such legislation is a major question.

Published online: 23 May 2023

References

1. *Nat. Mach. Intell.* **5**, 1 (2023).
2. Weidinger, L. et al. *FAccT '22: 2022 ACM Conference on Fairness, Accountability, and Transparency* 214–229 (ACM, 2022).
3. Ji, Z. et al. *ACM Comput. Surv.* **55**, 248 (2023).