

seeing no external submissions to an open database on water purification that was created more than one year ago⁴.

The delegates assembling in New York need to accept that their countries' visions will not be realized until all nations can somehow carve out a path to cooperate amid tension and conflict. Research can help to provide at least some of the right language, which is why it needs to be taken on board when decisions are being made. We in the Nature Portfolio intend to play our fullest part to make that happen.

1. *Nature Water* 1, 1 (2023).
2. Michalak, A. M. et al. *Nature Water* 1, 10–18 (2023).
3. Xu, R. et al. *Nature Water* 1, 113–122 (2023).
4. Verbeke, R. *Nature Water* 1, 7–9 (2023).

Tools such as ChatGPT threaten transparent science; here are our ground rules for their use

As researchers dive into the brave new world of advanced AI chatbots, publishers need to acknowledge the tools' legitimate uses and lay down clear guidelines to avoid abuse.

It has been clear for several years that artificial intelligence (AI) is gaining the ability to generate fluent language, churning out sentences that are increasingly hard to distinguish from text written by people. Last year, *Nature* reported that some scientists were already using chatbots as research assistants – to help organize their thinking, generate feedback on their work, assist with writing code and summarize research literature (*Nature* **611**, 192–193; 2022).

But the release of the AI chatbot ChatGPT in November has brought the capabilities of such tools, known as large language models (LLMs), to a mass audience. Its developers, OpenAI in San Francisco, California, have made the chatbot free to use and easily accessible for people who don't have technical expertise. Millions are using it, and the result has been an explosion of fun and sometimes frightening writing experiments that have turbocharged the growing excitement and consternation about these tools.

ChatGPT can write presentable student essays, summarize research papers, answer questions well enough to pass medical exams and generate helpful computer code. It has produced research abstracts good enough that scientists found it hard to spot that a computer had written them. Worryingly for society, it could also make

 No LLM tool will be accepted as a credited author on a research paper.”

spam, ransomware and other malicious outputs easier to produce. Although OpenAI has tried to put guard rails on what the chatbot will do, users are already finding ways around them.

The big worry in the research community is that students and scientists could deceitfully pass off LLM-written text as their own, or use LLMs in a simplistic fashion (such as to conduct an incomplete literature review) and produce work that is unreliable. Several preprints and published articles have already credited ChatGPT with formal authorship.

That's why it is high time researchers and publishers laid down ground rules about using LLMs ethically. *Nature*, along with all Springer Nature journals, has formulated the following two principles, which have been added to our existing guide to authors (see go.nature.com/3j1jxsw). As *Nature's* news team has reported, other scientific publishers are likely to adopt a similar stance (see page 620).

First, no LLM tool will be accepted as a credited author on a research paper. That is because any attribution of authorship carries with it accountability for the work, and AI tools cannot take such responsibility.

Second, researchers using LLM tools should document this use in the methods or acknowledgements sections. If a paper does not include these sections, the introduction or another appropriate section can be used to document the use of the LLM.

Pattern recognition

Can editors and publishers detect text generated by LLMs? Right now, the answer is 'perhaps'. ChatGPT's raw output is detectable on careful inspection, particularly when more than a few paragraphs are involved and the subject relates to scientific work. This is because LLMs produce patterns of words based on statistical associations in their training data and the prompts that they see, meaning that their output can appear bland and generic, or contain simple errors. Moreover, they cannot yet cite sources to document their outputs.

But in future, AI researchers might be able to get around these problems – there are already some experiments linking chatbots to source-citing tools, for instance, and others training the chatbots on specialized scientific texts.

Some tools promise to spot LLM-generated output, and *Nature's* publisher, Springer Nature, is among those developing technologies to do this. But LLMs will improve, and quickly. There are hopes that creators of LLMs will be able to watermark their tools' outputs in some way, although even this might not be technically foolproof.

From its earliest times, science has operated by being open and transparent about methods and evidence, regardless of which technology has been in vogue. Researchers should ask themselves how the transparency and trustworthiness that the process of generating knowledge relies on can be maintained if they or their colleagues use software that works in a fundamentally opaque manner.

That is why *Nature* is setting out these principles: ultimately, research must have transparency in methods, and integrity and truth from authors. This is, after all, the foundation that science relies on to advance.