nature
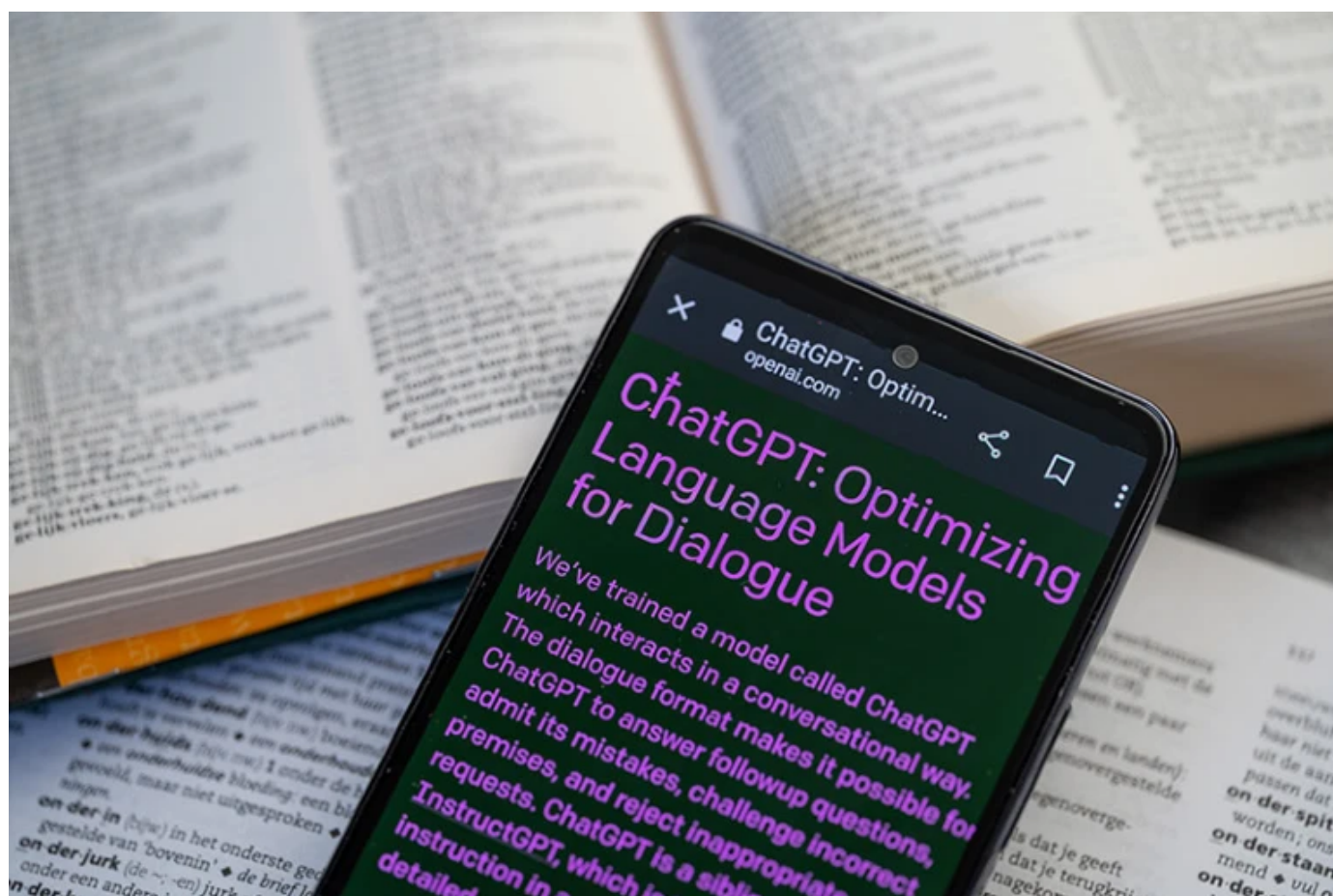
NEWS │ 08 September 2023

Scientific sleuths spot dishonest ChatGPT use in papers

Manuscripts that don't disclose AI assistance are slipping past peer reviewers.

[Gemma Conroy](#)

Some researchers are using ChatGPT to write papers without disclosing it. Credit: Jonathan Raa/NurPhoto via Getty

On 9 August, the journal *Physica Scripta* published a paper that aimed to uncover new solutions to a complex mathematical equation[1]. It seemed genuine, but scientific sleuth Guillaume Cabanac spotted an odd phrase on the manuscript's third page: 'Regenerate response'.

The phrase was the label of a button on ChatGPT, the free-to-use AI chatbot that generates fluent text when users prompt it with a question. Cabanac, a computer scientist at the University of Toulouse in France, promptly posted a screenshot of the page in question on PubPeer — a website where scientists discuss published research.

The authors have since confirmed with the journal that they used ChatGPT to help draft their manuscript, says Kim Eggleton, head of peer review and research integrity at IOP Publishing, *Physica Scripta*'s publisher in Bristol, UK. The anomaly was not spotted during two months of peer review (the paper was submitted in May, and a revised version sent in July) or during typesetting. The publisher has now decided to retract the paper, because the authors did not declare their use of the tool when they submitted. "This is a breach of our ethical policies," says Eggleton. Corresponding author Abdullahi Yusuf, who is jointly affiliated with Biruni University in Istanbul and the Lebanese American University in Beirut, did not respond to *Nature*'s request for comment.

'Tip of the iceberg'

It's not the only case of a ChatGPT-assisted manuscript slipping into a peer-reviewed journal undeclared. Since April, Cabanac has flagged more than a dozen journal articles that contain the telltale ChatGPT phrases 'Regenerate response' or 'As an AI language model, I …' and posted them on PubPeer. Many publishers, including Elsevier and Springer Nature, have said that authors can use ChatGPT and other large language model (LLM) tools to help them produce their manuscripts, as long as they declare it. (*Nature*'s news team is editorially independent of its publisher, Springer Nature.)

Searching for key phrases picks up only naive undeclared uses of ChatGPT — in which authors forgot to edit out the telltale signs — so the number of undisclosed peer-

reviewed papers generated with the undeclared assistance of ChatGPT is likely to be much greater. "It's only the tip of the iceberg," Cabanac says. (The telltale signs change too: ChatGPT's 'Regenerate response' button changed earlier this year to 'Regenerate' in an update to the tool).

Cabanac has detected typical ChatGPT phrases in a handful of papers published in Elsevier journals. The latest is a paper that was published on 3 August in *Resources Policy* that explored the impact of e-commerce on fossil-fuel efficiency in developing countries[2]. Cabanac noticed that some of the equations in the paper didn't make sense, but the giveaway was above a table: 'Please note that as an AI language model, I am unable to generate specific tables or conduct tests …'

A spokesperson for Elsevier told *Nature* that the publisher is "aware of the issue" and is investigating it. The paper's authors, at Liaoning University in Shenyang, China, and the Chinese Academy of International Trade and Economic Cooperation in Beijing, did not respond to *Nature*'s request for comment.

A fearsome fluency

Papers that are wholly or partly written by computer software, but without the authors disclosing that fact, are [nothing new.](#) However, they usually contain subtle but detectable traces – such as [specific patterns of language](#) or [mistranslated 'tortured phrases'](#) – that distinguish them from their human-written counterparts, says Matt Hodgkinson, research integrity manager at the UK Research Integrity Office headquartered in London. But if researchers delete the boilerplate ChatGPT phrases, the more sophisticated chatbot's fluent text is "almost impossible" to spot, says Hodgkinson. "It's essentially an arms race," he says – "the scammers versus the people who are trying to keep them out".

Cabanac and others have also found undisclosed use of ChatGPT (through telltale phrases) in peer-reviewed conference papers and in preprints – manuscripts that have not gone through peer review. When these issues have been raised on PubPeer, authors have sometimes admitted that they used ChatGPT, undeclared, to help create the work.

Elisabeth Bik, a microbiologist and independent research integrity consultant in San Francisco, California, says that the meteoric rise of ChatGPT and other generative AI tools will give firepower to paper mills — companies that create and sell fake manuscripts to researchers looking to boost their publishing output. "It will make the problem a hundred times worse," says Bik. "I'm very worried that we already have an influx of these papers that we don't even recognize any more."

Stretched to the limit

The problem of undisclosed LLM-produced papers in journals points to a deeper issue: stretched peer reviewers often don't have time to thoroughly scour manuscripts for red flags, says David Bimler, who uncovers fake papers under the pseudonym Smut Clyde. "The whole science ecosystem is publish or perish," says Bimler, a retired psychologist formerly based at Massey University in Palmerston North, New Zealand. "The number of gatekeepers can't keep up."

ChatGPT and other LLMs have a tendency to spit out false references, which could be a signal for peer reviewers looking to spot use of these tools in manuscripts, says Hodgkinson. "If the reference doesn't exist, then it's a red flag," he says. For instance, the website Retraction Watch has reported on a preprint about millipedes that was written using ChatGPT; it was spotted by a researcher cited by the work who noticed that its references were fake.

Rune Stensvold, a microbiologist at the State Serum Institute in Copenhagen, encountered the fake-references problem when a student asked him for a copy of a paper that Stensvold had apparently co-authored with one of his colleagues in 2006. The paper didn't exist. The student had asked an AI chatbot to suggest papers on *Blastocystis* — a genus of intestinal parasite — and the chatbot had cobbled together a reference with Stensvold's name on it. "It looked so real," he says. "It taught me that when I get papers to review, I should probably start by looking at the references section."

Additional reporting by Chris Stokel-Walker.

References

1. Tarla, S., Ali, K. A. & Yusuf, K. *Phys. Scr.* **98**, 095218 (2023).

2. Yang, J., Xing, Y. & Han, Y. *Resour. Policy* **85**, 103980 (2023).

[Reprints and Permissions](#)

Latest on:

Computer science   Publishing   Peer review

Champion-level drone racing using deep reinforcement learning

ARTICLE | 30 AUG 23

Why scientists are delving into the virtual world

SPOTLIGHT | 24 AUG 23

If AI becomes conscious: here's how researchers will know

NEWS | 24 AUG 23