

ChatGPT writes convincing fake scientific abstracts that fool reviewers in study

January 13 2023, by Marla Paul



ItBlinded human reviewers—when given a mix real and falsely generated abstracts—could only spot ChatGPT generated abstracts 68% of the time. The reviewers also incorrectly identified 14% of real abstracts as being AI generated. Credit: Northwestern University

Could the new and wildly popular chatbot ChatGPT convincingly

produce fake abstracts that fool scientists into thinking those studies are the real thing?

That was the question worrying Northwestern Medicine physician-scientist Dr. Catherine Gao when she designed a study—collaborating with University of Chicago scientists—to test that theory.

Yes, scientists can be fooled, their new study reports. Blinded human reviewers—when given a mix real and falsely generated abstracts—could only spot ChatGPT generated abstracts 68% of the time. The reviewers also incorrectly identified 14% of real abstracts as being AI generated.

"Our reviewers knew that some of the abstracts they were being given were fake, so they were very suspicious," said corresponding author Gao, an instructor in pulmonary and critical care medicine at Northwestern University Feinberg School of Medicine. "This is not someone reading an abstract in the wild. The fact that our reviewers still missed the AI-generated ones 32% of the time means these abstracts are really good. I suspect that if someone just came across one of these generated abstracts, they wouldn't necessarily be able to identify it as being written by AI."

The hard-to-detect fake abstracts could undermine science, Gao said. "This is concerning because ChatGPT could be used by 'paper mills' to fabricate convincing scientific abstracts," Gao said. "And if other people try to build their science off these incorrect studies, that can be really dangerous."

Paper mills are illegal organizations that produce fabricated scientific work for profit.

The ease with which ChatGPT produces realistic and convincing

abstracts could increase production by paper mills and fake submissions to journals and scientific conferences, Gao worries.

AI sleuths can identify AI fakes

For the study, Gao and co-investigators took titles from recent papers from high-impact journals and asked ChatGPT to generate abstracts based on that prompt. They ran these generated abstracts and the original abstracts through a plagiarism detector and AI output detector, and had blinded human reviewers try to differentiate between generated and original abstracts. Each reviewer was given 25 abstracts that were a mixture of the generated and original abstracts and asked to give a binary score of what they thought the abstract was.

"The ChatGPT-generated abstracts were very convincing," Gao said, "because it even knows how large the patient cohort should be when it invents numbers." For a study on hypertension, which is common, ChatGPT included tens of thousands of patients in the cohort, while a study on a monkeypox had a much smaller number of participants.

"Our reviewers commented that it was surprisingly difficult to differentiate between the real and fake abstracts," Gao said.

The study found that the fake abstracts did not set off alarms using traditional plagiarism-detection tools. However, in the study, AI output detectors such as GPT-2 Output Detector, which is available online and free, could discriminate between real and fake abstracts.

"We found that an AI output detector was pretty good at detecting output from ChatGPT and suggest that it be included in the scientific editorial process as a screening process to protect from targeting by organizations such as paper mills that may try to submit purely generated data," Gao said.

ChatGPT also can be used for good

But ChatGPT can also be used for good, said senior study author Yuan Luo, director of the Institute for Augmented Intelligence in Medicine at Feinberg.

"AI language models such as ChatGPT have a potential to help automate the writing process, which is often the speed bottleneck in knowledge generation and dissemination," Luo said. "The results from the paper showed this is likely doable for the field of medicine, but we need to bridge certain ethical and practical gaps."

For example, is AI-assisted writing still considered original, Luo asked. Also, AI-generated text currently has difficulty in proper citation, which is a must for scientific writing, he noted.

"Generative text technology has a great potential for democratizing science, for example making it easier for non-English-speaking scientists to share their work with the broader community," said senior author Dr. Alexander Pearson, director of data sciences and the Head/Neck Cancer Program in Hematology/Oncology at the University of Chicago. "At the same time, it's imperative that we think carefully on best practices for use."

The research is available on the *bioRxiv* preprint server.

More information: Catherine A. Gao et al, Comparing scientific abstracts generated by ChatGPT to original abstracts using an artificial intelligence output detector, plagiarism detector, and blinded human reviewers, *bioRxiv* (2022). [DOI: 10.1101/2022.12.23.521610](https://doi.org/10.1101/2022.12.23.521610)

Provided by Northwestern University

Citation: ChatGPT writes convincing fake scientific abstracts that fool reviewers in study (2023, January 13) retrieved 25 April 2024 from <https://techxplore.com/news/2023-01-chatgpt-convincing-fake-scientific-abstracts.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.