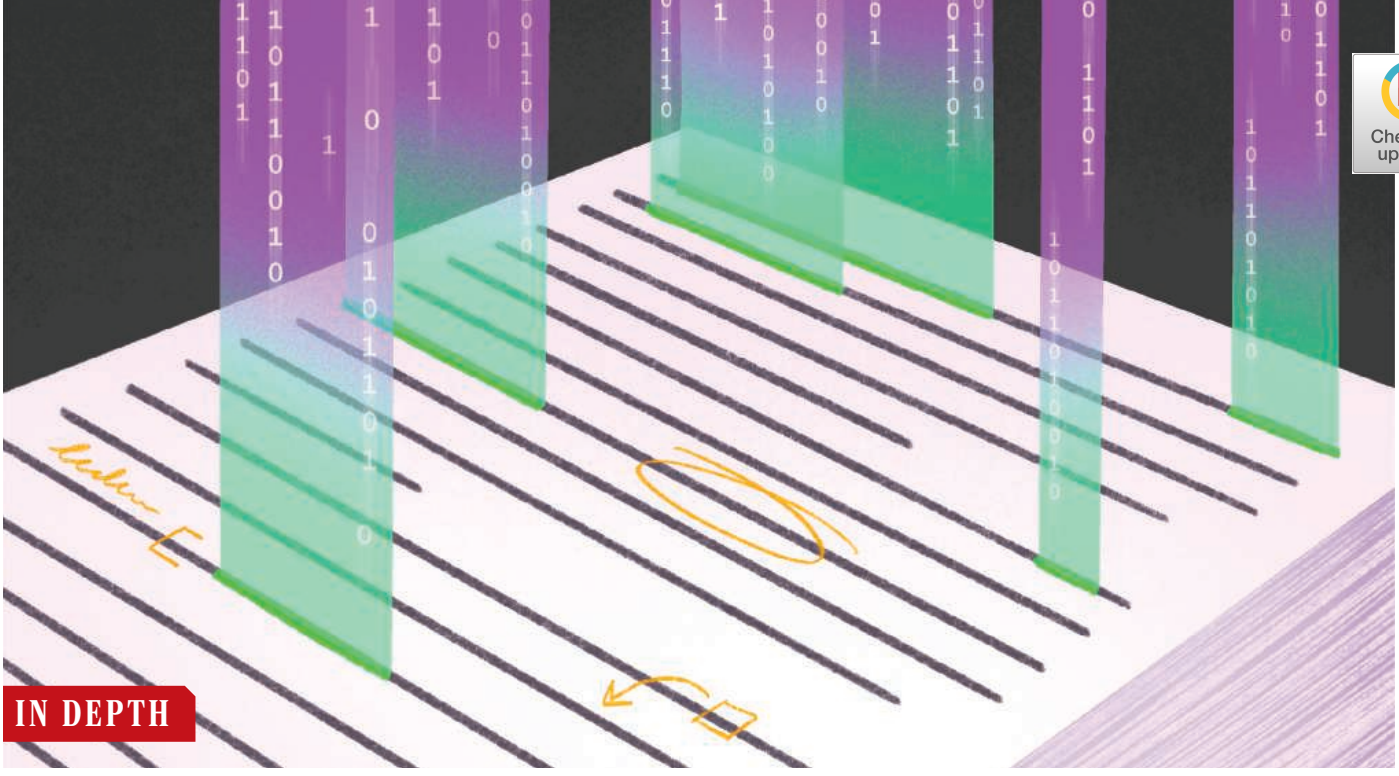

IN DEPTH

PUBLISHING

Journals take up arms against AI-written text

Many ask authors to disclose use of ChatGPT and other generative artificial intelligence

By **Jeffrey Brainard**

"It's all we've been talking about since November," says Patrick Franzen, publishing director for SPIE, the international society for optics and photonics. He's referring to ChatGPT, the artificial intelligence (AI)-powered chatbot unveiled that month. In response to a prompt, ChatGPT can spin out fluent and seemingly well-informed reports, essays—and scientific manuscripts. Worried about the ethics and accuracy of such content, Franzen and managers at other journals are scrambling to protect the scholarly literature from a potential flood of manuscripts written in whole or part by computer programs.

Some publishers have not yet formulated policies. Most of those that have avoid an outright ban on AI-generated text, but ask authors to disclose their use of the automated tools, as SPIE is likely to do. For now, editors and peer reviewers have few alternatives, as they lack enforcement tools. No software so far can consistently detect the synthetic text the majority of the time.

When the online tool ChatGPT was made available for free public use, scientists were among those who flocked to try it out. (ChatGPT's creator, the U.S.-based company OpenAI, has since limited access to subscribers.) Many reported its unprecedented and uncanny ability to create plausible-sounding text, dense with seemingly factual detail. ChatGPT and its brethren—including Google's Bard, unveiled earlier this month for select users, and Meta's Galactica, which was briefly available for public use in November 2022—are AI algorithms called large language models, trained on vast numbers of text samples pulled from the internet. The software identifies patterns and relationships among words, which allows the models to generate relevant responses to questions and prompts.

In some cases, the resulting text is indistinguishable from what people would write. For example, researchers who read medical journal abstracts generated by ChatGPT failed to identify one-third of them as written by machine, according to a December 2022 preprint. AI developers are expected to create even more powerful versions, including ones trained specifically on scientific literature—a prospect that has sent a shock wave through the scholarly publishing industry.

So far, scientists report playing around with ChatGPT to explore its capabilities, and a few have listed ChatGPT as a co-author on manuscripts. Publishing experts worry such limited use could morph into a spike of manuscripts containing substantial chunks of AI-written text.

One concern for journal managers is accuracy. If the software hasn't been exposed to enough training data to generate a correct response, it will often fabricate an answer, computer scientists have found. In November, Meta took down the public interface for Galactica, its scientist-specific large language model, just days after it was unveiled—users had identified myriad factual errors in the generated text. And a 2022 preprint study of Sparrow, an information-seeking chatbot developed by a Google subsidiary, found that up to 20% of its responses contained errors. AI text may also be biased toward established scientific ideas and hypotheses contained in the content on which the algorithms were trained. Journal editors also worry about ethics, suggesting authors who use text generators are sometimes presenting the outputs as if they wrote them—a transgression others have dubbed "aigiarism."

Many journals' new policies require that authors disclose use of text-generating tools and ban listing a large language model such as ChatGPT as a co-author, to underscore the human author's responsibility for ensuring the text's accuracy. That is the case for *Nature* and all Springer Nature journals, the JAMA Network, and groups that advise on best practices in publishing, such as the Committee on Publication Ethics and the World Association of Medical Editors. But at least one publisher has taken a tougher line:

ILLUSTRATION: A. MASTIN/*SCIENCE*

Downloaded from https://www.science.org at University of Massachusetts Boston on November 20, 2023

Corrected 24 February 2023. See full text.

The *Science* family of journals announced a complete ban on generated text last month. The journals may loosen the policy in the future depending on what the scientific community decides is acceptable use of the text generators, Editor-in-Chief Holden Thorp says. "It's a lot easier to loosen our criteria than it is to tighten them."

Some publishing officials are still working out the details, such as when they might ask journal staff, editors, or reviewers to examine or fact check generated text disclosed by authors—tasks that would add to what is often already a heavy volunteer workload. Editors at the Taylor & Francis publishing group, whose forthcoming rule will likely require disclosure of such text, might sometimes ask authors to specify what parts of their manuscript were written by a computer, says Sabina Alam, director of publishing ethics and integrity. Searching for papers to include in a systematic review may be a legitimate use if the researcher follows proper methods in deciding which papers to include, for example, she says, whereas cutting and pasting it into a perspective or opinion piece "is not OK because it's not your perspective." The policy will likely evolve as the publishing industry gets more experience working with such manuscripts, she says. "We're seeing this as a phased approach. It's really early days."

Journal managers also say they are hoping to monitor the new technology using more technology: automated detectors that can flag synthetic text. But that isn't easy, says Domenic Rosati, a senior research scientist at scite.ai, a company that develops software to assist scientific publishers. "We're well past the time, in science in particular, of being able to say it's obvious [certain text came from] a machine because of its fluency or its lack of truthfulness."

Current detectors leave much to be desired. OpenAI unveiled its "classifier" last month, which categorizes submitted text on a scale from "likely" to have been written by a computer to "very unlikely." The classifier was training using paired samples of human writing and computer-generated text from 34 algorithms from five different companies, including OpenAI itself. But OpenAI concedes several limitations. The tool, still under development, correctly applies the "likely" label only 26% of the time. People can fool it by editing computer-generated text. And it may not consistently identify synthetic text on topics that were not included in the training data. Computer scientists say these limitations typically apply to other detectors as well.

Better solutions may be on the horizon. OpenAI said in December 2022 it is working on ways to "watermark" the generated text. It would program its models to insert words, spelling, and punctuation in a telltale order to create a secret code detectable by search engines. And last month, a team at Stanford University published a preprint describing DetectGPT, which, unlike other detectors, doesn't require training. This algorithm examines text by creating multiple, random variations and querying a text generator to rank which versions it prefers. The extent to which the generator the team studied—developed by OpenAI and similar to ChatGPT—prefers the original text versus the altered versions is consistently different for human-written versus AI-generated text, allowing DetectGPT to predict the likelihood that a sample came from a particular machine. But DetectGPT needs more development before journal editors can exclusively rely on its results to make decisions on manuscripts, for example, says Eric Mitchell, a doctoral student who led the Stanford team. The company TurnItIn, which markets a widely used plagiarism detector, said last week it plans to roll out a synthetic text detector as early as April. TurnItIn says the tool, trained on academic writing, can identify 97% of text generated by ChatGPT, with a false positive rate of one in 100.

> **"We're seeing this as a phased approach. It's really early days."**
> **Sabina Alam**, Taylor & Francis

A further computational challenge is rating the factual accuracy of robot-generated text and the quality of its summaries, Rosati says. His firm is working on an automated checker that would scour existing scientific literature to determine whether a particular citation in a manuscript actually presents the finding the manuscript claims it does, for example. That tool could turn up references fabricated by a machine as well as irrelevant or incorrect ones provided by humans, Rosati says.

Some editors see promise as well as peril in the emergence of ChatGPT and its relatives. Programmers are developing and refining software that creates hypotheses, writes computer code, and analyzes data, for example, to make researchers' work more efficient and productive. "I can see scenarios where there will be causes for concern, but then I also see a tremendous amount of opportunity with these types of tools," says Magdalena Skipper, editor-in-chief of *Nature*. "As with every tool, we have to understand the limitations. … It demands that we pay close attention to how these tools are developed, how they are used, and in what context." ■

MARINE SCIENCE

Iron stress threatens Southern Ocean phytoplankton

Lack of the nutrient limits the plants' productivity, key to climate and ecosystems

By **Warren Cornwall**

Off the icy shores of Antarctica each spring, an explosion of life unfolds that is so large it's visible from space. As iron-rich waters rise from below, the surface of the Southern Ocean swirls with psychedelic clouds of bright green phytoplankton—single-celled plants that suck up carbon from the atmosphere and form the base of the food chain by sustaining krill, which is in turn a major food source for fish, whales, and penguins.

Now, a group of scientists says that over the past quarter-century this seasonal bloom, a critical player in ecosystems and climate, might be at risk. Phytoplankton across the Southern Ocean are increasingly starved of iron—a building block for their photosynthetic machinery—and there are signs their productivity might be declining. The discovery, published this week in *Science* (p. 834), is a surprise, directly counter to the surge of productivity that many climate models predicted for the coming century.

The apparent speed of the change "is really alarming," says Adrian Marchetti, a biological oceanographer at the University of North Carolina, Chapel Hill, who studies phytoplankton but was not directly involved in the research. A big drop in phytoplankton "could really affect the global carbon cycle," adds Alison Gray, a University of Washington, Seattle, oceanographer who studies ocean carbon.

Ocean iron levels, although known to be an important factor limiting phytoplankton productivity in the Southern Ocean, are notoriously difficult to study. Neither robotic sensors nor research ships routinely look for the nutrient. So scientists have recently begun to infer its levels by looking for signals that phytoplankton are coping with an iron shortage.

Downloaded from https://www.science.org at University of Massachusetts Boston on November 20, 2023

Corrected 24 February 2023. See full text.

Journals take up arms against AI-written text

Jeffrey Brainard

Science **379** (6634), . DOI: 10.1126/science.adh2762

View the article online
https://www.science.org/doi/10.1126/science.adh2762
Permissions
https://www.science.org/help/reprints-and-permissions

Downloaded from https://www.science.org at University of Massachusetts Boston on November 20, 2023

Use of this article is subject to the Terms of service

Science (ISSN 1095-9203) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.

Copyright © 2023 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works