PUBLISHING

Far more authors use AI than admit it

Study by cancer publisher highlights promise of AI detectors and the challenges they pose

JEFFREY BRAINARD

After ChatGPT debuted in late 2022 and wowed users with its humanlike fluency, many academic journals rolled out policies requiring authors to disclose whether they had employed artificial intelligence (AI) to help write their papers. But new evidence from one publisher suggests four times as many authors use AI as admit to it—and that peer reviewers are turning to it, too, even though they are asked not to.

The new study, run by the American Association for Cancer Research (AACR), investigated the 10 journals the society publishes. AACR launched it after some authors wondered whether the peer-review reports on papers they had submitted were AI-generated, says Daniel Evanko, who oversees AACR's editorial systems. It made use of a recently developed AI detector the AACR team and others say appears to be highly accurate.

From 1 January to 30 June, the team found, 36% of the abstracts in 7177 manuscripts submitted to AACR contained at least some AI-generated text. But when asked in an automatic step in the submission process to disclose any use of AI to prepare the manuscript, authors only did so for 9% of the papers studied.

Earlier studies tried to quantify the use of AI in papers and peer reviews. But the new study, presented at the International Congress on Peer Review and Scientific Publication earlier this month, is one of the first to assess the reliability of author disclosures. "Disclosures on their own have virtually no value without some means of determining their accuracy," Evanko says.

The work is "a good place to start" to address the problem, says Roy Perlis, editor-in-chief of the JAMA Network's content channel JAMA+AI and a psychiatrist at Massachusetts General Hospital. But AI detectors produce false positives, and human editors must use judgment in interpreting their readings, he says. "There is a real risk that we plug these things into our [editorial] pipelines and treat their outputs as if they are infallible."

Evanko says he was initially "extremely skeptical" about the high accuracy claims from the new detector his team ultimately used in its study, the AI Detection Dashboard from Pangram Labs. Pangram's detector, unveiled in 2024, relies on a form of AI called deep learning, a computational method also used in

large language models (LLMs) such as ChatGPT. In a preprint that year, its creators described the detector's text classifier as more accurate than others because they trained it using an unusual method and data set.

They started with a large body of human-written texts. They then prompted LLMs to produce a similar version of each text that matched its style, tone, and semantic content. They trained their text classifier to spot telltale differences between the two, progressively modifying the prompts so the LLMs generated text increasingly difficult for their classifier to distinguish from human-written text. The tool produces scores on a 10-point scale reflecting the likelihood of AI use.

Despite Evanko's initial misgivings, he was reassured that the tool is unusually good at avoiding false positives when he ran it on AACR's submissions from 2020 and 2021, before ChatGPT. It flagged well under 1% of those manuscripts as possibly AI-generated.

After ChatGPT arrived, the detector showed, AI-generated text steadily became more common in AACR papers' abstracts, methods sections, and peer-review reports. (Evanko's study only covered those kinds of texts because AACR's database includes them in a format that is readily analyzable.) In addition to the high proportion of abstracts with AI-generated text, Evanko's team found it in nearly 15% of the methods sections and 7% of reviewer reports in the last quarter of 2024.

He speculates authors are not disclosing AI use because they fear journals will reject their manuscript, even if AI was used only for editing the manuscript and not generating text. Some evidence supports that reviewers penalize this use. But how authors perceive that risk varies by field, according to a 2024 survey of more than 800 researchers, co-authored by Amy Zhang of the University of Washington, who studies human-computer interaction. Respondents in computer science were more likely than those in biology and medicine to say they were comfortable with disclosure. In computer science, "it just has become so common and normal to use," she says. But norms about AI use "are unsettled in these other fields." The International Association of Scientific, Technical & Medical Publishers

reported in April that many authors are confused about when they should report AI use; the group, known as STM, expects to finalize updated guidelines this week.

Using AI for editing manuscripts and other purposes can be legitimate, Evanko and Perlis say, especially when authors are not native speakers of English. In fact, Evanko found that manuscripts from countries where English is not an official language were flagged for AI-generated text twice as often as those from English-speaking countries, perhaps because authors turned to AI to improve their writing. But AI-generated text can also be one of many markers of sub-

missions that might have come from paper mills, Evanko adds. A spate of letters to the editor recently submitted to an AACR journal was all generated by the DeepSeek LLM, Pangram's tool indicated.

AACR is considering next steps in response to Evanko's findings, including using the new tool to screen all submissions. But with more than 2500 AACR submissions flagged for AI-generated abstracts from January to June alone, "It's too many to put a human in the loop" to follow up on each undisclosed instance, he says. The publisher might start by sending automated emails to authors asking for an

explanation, as it does about other deficiencies in manuscripts.

But Perlis says he's not persuaded that AI-text detectors are accurate enough to help publishers and editors deal with the machine-generated text appropriately. He wants common performance benchmarks and more data about how the detectors perform on manuscripts from different fields of science before they are used routinely. "We want to encourage people to continue to develop these kinds of tools," he says. "We also want to acknowledge that there will absolutely be an arms race—the better the tools get, the harder people will work to circumvent them." □

IN FOCUS
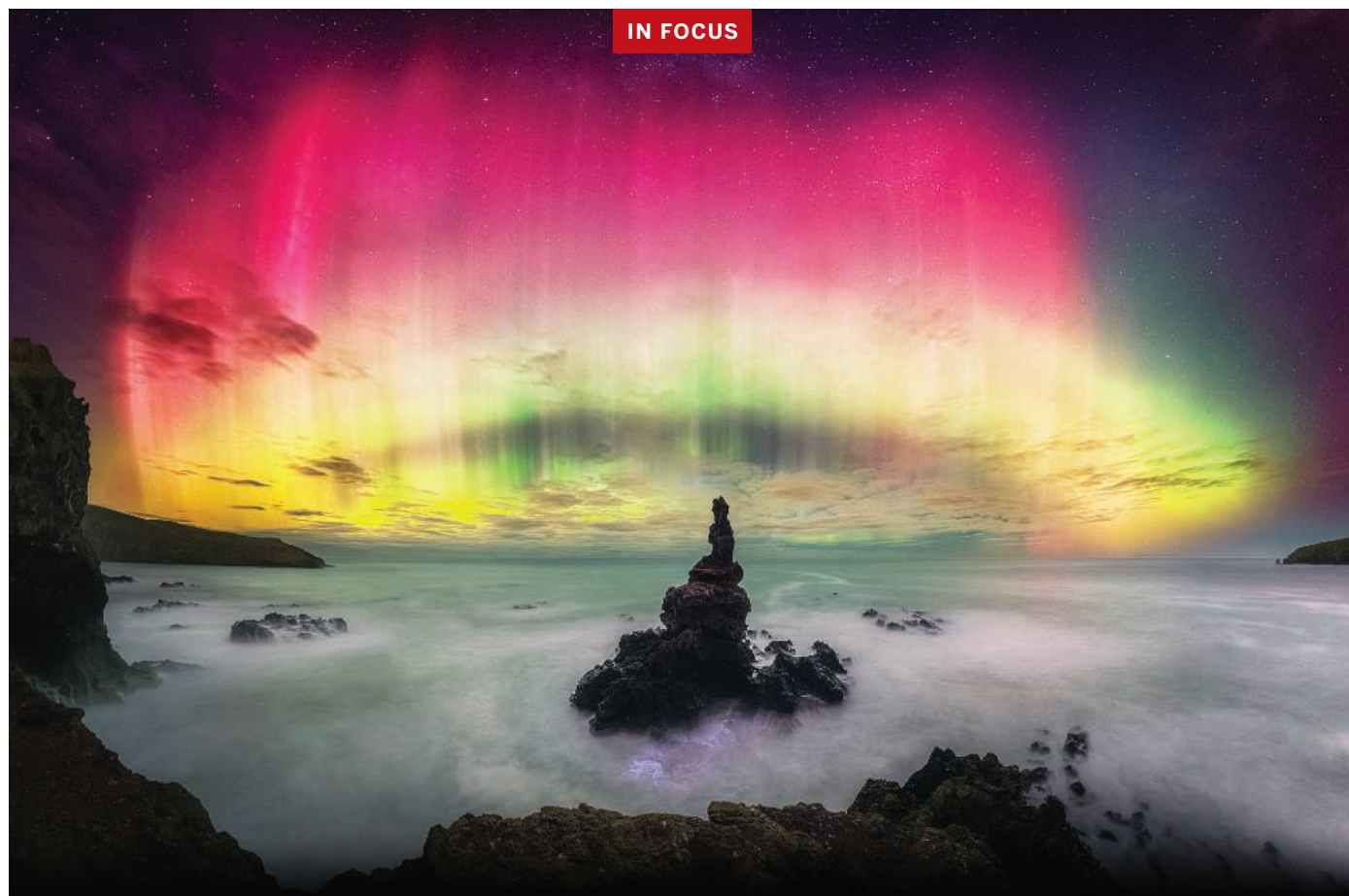
A prize-winning light show

This image of the southern lights, or the aurora australis, captured in Tumbledown Bay, New Zealand, in May 2024, was among the winners of this year's Astronomy Photographer of the Year competition, run by the Royal Observatory Greenwich. That month, scientists recorded the strongest geomagnetic storm in more than 2 decades, brought on as blobs of plasma ejected from the Sun slammed into Earth's magnetic field. The huge amounts of energy released led to aurorae that lit up night skies around the globe at lower latitudes than usual. Of the New Zealand display, photographer Kavan Chay noted, "The reds were a level of intensity I had never experienced." —Katie Langin