

THE GROWING PROBLEM OF HALLUCINATED CITATIONS

A *Nature* analysis suggests that tens of thousands of publications from 2025 might include invalid references generated by AI. By **Miryam Naddaf** and **Elizabeth Quill**

Earlier this year, computer scientist Guillaume Cabanac received a notification from Google Scholar that one of his publications had been cited in a paper published in the *International Dental Journal*¹. That was unexpected, because his research on spotting fabricated papers doesn't typically intersect with dentistry. "I was very surprised to see that I couldn't recognize my own reference," says Cabanac, who is based at the University of Toulouse in France.

The title in the citation resembled that of a preprint² he had posted in 2021 and never published formally, but the journal was listed as *Nature* and the DOI – the unique identifier assigned by publishers and preprint repositories – did not lead to the original preprint. "I got very concerned," adds Cabanac, who immediately suspected that the citation had been hallucinated by artificial intelligence.

This is just one example of a rapidly growing problem. Surveys and related studies have



shown that researchers are increasingly using large language models (LLMs) to help to conduct literature searches, write manuscripts and format bibliographies. And sometimes, these models generate non-existent academic references.

Over the past year, efforts have begun turning up such hallucinated citations in the literature. One analysis of nearly 18,000 papers accepted by three computer-science conferences found a sharp increase in references that cannot be traced to actual scholarly publications³. The results, reported in January, indicated that 2.6% of papers in 2025 had a least one potentially hallucinated citation – up from about 0.3% in 2024. Another analysis, released in February, estimated that 2–6% of papers in four other 2025 computer-science conferences included references with rephrased titles or citations of publications that the authors couldn't verify by searching through databases and journal archives⁴.

And although the scale of the problem

remains uncertain, it's clear that not only conferences are affected. An exclusive analysis conducted by *Nature's* news team, in collaboration with Grounded AI, a company based in Stevenage, UK, suggests that at least tens of thousands of 2025 publications, including journal papers and books, as well as conference proceedings, probably contain invalid references generated by AI.

Grounded AI is among the companies offering publishers tools for screening submissions for problematic references. Several publishers told *Nature* reporters that they have been exploring such tools or developing in-house versions.

But some researchers are concerned that the problem will soon get out of hand. "We're going to see a flood of fake references," says Alison Johnston, a political scientist at Oregon State University in Corvallis.

Another issue is deciding what to do about hallucinated citations that make it into the published literature. That's a problem that academic publishers are wrestling with right now.

Sources of error

Citation errors are not new to academic publishing. "Even before generative AI, we already had so many inaccuracies in citations," says Mohammad Hosseini, who studies research ethics and integrity at Northwestern University Feinberg School of Medicine in Chicago, Illinois. Issues have tended to include misspelling of authors' names or errors in the year of publication, the title of the journal or the DOI. Another issue has been discrepancies between the information in the cited work and the details given by the paper citing it^{5,6}.

"Now the problem is not just inaccuracy, it's about fake citations. It's about fabricated citations, which is a whole different problem," says Hosseini.

Publishers told *Nature* that they are seeing increases in the number of fabricated and inaccurate citations in submissions, and they are taking steps to tackle the issue.

Johnston, co-lead editor of the *Review of International Political Economy (RIPE)*, a journal published by the UK-based Taylor & Francis, says that she rejected 25% of some 100 submissions in January "because of fake references". She uses the plagiarism-detection software iThenticate to flag unusual or partial matches between the references in submitted papers and published bibliographies. Then she manually checks the suspicious citations. "I'm doing things now to try and detect hallucinated references that I wasn't doing prior to 2025," she says.

Frontiers, based in Lausanne, Switzerland, has developed an in-house AI tool for flagging integrity issues at the point of submission, including references to irrelevant or retracted work and hallucinated citations. "Around 5% [of manuscripts] show potential reference-related

issues flagged through our checks," says Elena Vicario, Frontiers' head of research integrity. But "not all flagged references ultimately turn out to be genuinely problematic", she adds. That makes it challenging, Vicario says, to come up with a precise measure of the prevalence of any of these types of citation issue.

Experiments using AI chatbots to generate papers have provided insights into how often LLMs produce citation errors and what types of error they tend to make. In one study, researchers prompted OpenAI's GPT-4o LLM to generate six literature reviews on three mental-health disorders, and analysed the 176 references in those synthetic reviews⁷. Under these experimental conditions, they found that nearly 20% were fabricated references and could not be linked to actual research. And 45% of the remaining references, which corresponded to genuine publications, contained errors, often incorrect or invalid DOIs.

In some cases, including in references in published articles, all of the component parts

"It looks real to a human being, but is not actually a reference to a real thing."

are made up, says Kathryn Weber-Boer, director of scientometrics at the London-based company Digital Science. (The firm is operated by the Holtzbrinck Publishing Group, which is the majority shareholder of Springer Nature, which publishes *Nature*. *Nature's* news team is editorially independent of its publisher.) AI also hallucinates DOIs, both in references that are otherwise genuine as well as in fabricated ones, she adds.

AI-generated references commonly combine fragments of genuine publications, say researchers who have studied the issue (see 'How fakes can look real'). Joe Shockman, co-founder and chief executive of Grounded AI, calls such references 'Frankenstein' citations, likening their assembly to that of the fictional monster. "It looks real to a human being, but is not actually a reference to a real thing," says Shockman, who is based in Ashland, Oregon.

Although some types of error seem to implicate AI, others are less clear-cut, say researchers. "In today's landscape, we have to recognize that there are human errors and there are machine errors, and those can often overlap," says Weber-Boer.

Published problems

How many hallucinated citations are showing up in published research remains difficult to discern. To get an estimate, *Nature's* news team joined forces with Grounded AI, which has developed an AI tool called Veracity that checks citations against scholarly databases and across the web, flagging ones that are



Feature

invalid, irrelevant or cite retracted work.

Nature and Grounded AI collaborated to analyse more than 4,000 publications from last year, covering five leading publishers: Elsevier, Sage, Springer Nature, Taylor & Francis and Wiley. Grounded AI randomly sampled these papers from Europe PMC – a repository of open-access biomedical research articles – and the bibliometric database Crossref, to include equal number of publications per month from each of the five publishers. The sample included published papers as well as book chapters and conference proceedings, and it cut across all subject areas in these publishers' portfolios.

Grounded AI's tool looks for an exact match to a reference or the closest match it can find. It then flags citations with major issues, such as mismatched titles or DOIs, missing authors and incorrect journals, as well as more-minor issues. Citations that pointed to papers that couldn't be found even though they should be easy to find – because the journal in question is indexed by scholarly databases, for example – were marked as especially problematic.

After running the publications through the tool, Grounded AI assigned a risk score to each of the published papers, on the basis of the number of references that had major issues and how likely those issues were to have been generated by AI. Grounded AI determined that likelihood using data gleaned from a separate analysis that used two AI models to generate 20,000 synthetic papers; this allowed the company to identify the most common types of citation error that AI makes.

Nature manually checked the 100 most suspicious publications and confirmed that 65 contained at least one invalid reference, meaning that it pointed to a publication that did not seem to exist (see 'Finding the fabrications'). But 22 of the 100 most-suspicious papers had references that did point to genuine publications.

For the remaining 13 papers, it was unclear whether all their citations pointed to existing research or not. These 13 papers included references to articles that were said to be published in regional journals in languages other than English, and references that had mismatches in metadata that looked like plausible human errors, for example.

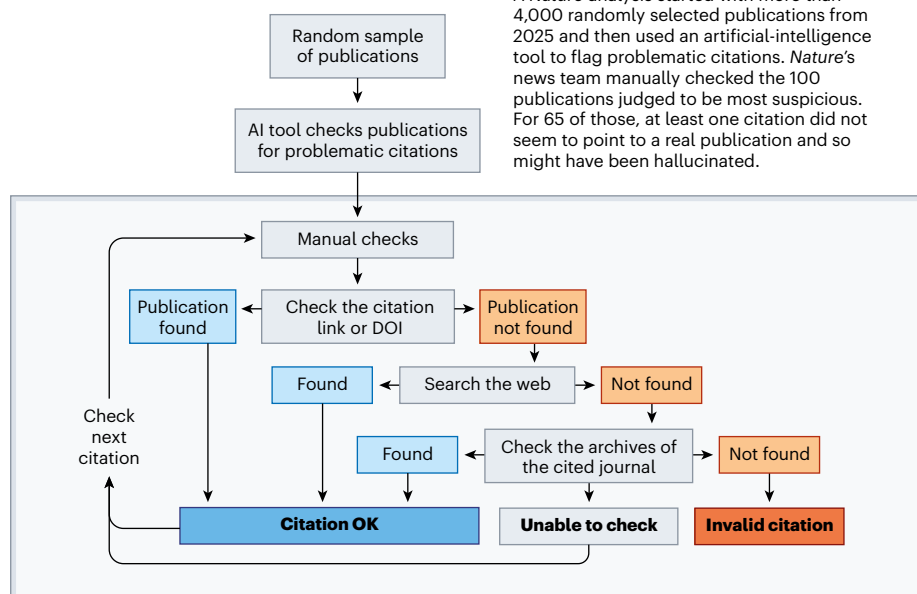
The analysis, which looked at reference lists from Crossref and full text from Europe PMC publications, turned up no clear trend across publishers. Each of the selected publishers had more than five publications with references that manual checks couldn't validate.

As a rough estimate, if the rate of 65 publications with at least one invalid reference out of some 4,000 publications analysed holds across the academic literature, it would suggest that more than 110,000 of the 7 million or so scholarly publications from 2025 contain invalid references.

Nick Morley, Grounded AI's co-founder and

FINDING THE FABRICATIONS

A *Nature* analysis started with more than 4,000 randomly selected publications from 2025 and then used an artificial-intelligence tool to flag problematic citations. *Nature*'s news team manually checked the 100 publications judged to be most suspicious. For 65 of those, at least one citation did not seem to point to a real publication and so might have been hallucinated.



chief product officer, says that the types of citation problem seen in 2025 are different from those found by his team before the proliferation of LLMs. This fact, he says, points to the use of AI as a leading culprit.

The true number of hallucinated references is almost certainly higher, says Weber-Boer, because the analysis focused on big publishers, which have more resources for checking citations systematically than do smaller publishers. Fields such as computer science, which has seen a surge in the use of LLMs to produce manuscripts⁸, might be more affected than other fields. What's more, the Grounded AI analysis turned up a few hundred more publications that had some risk of hallucinated citations,

"We know it's a problem, we just don't know how big the problem is."

suggesting that extra manual checking would have brought more such citations to light.

Spokespeople for all five publishers said that they check references as part of their screening and editing process, and they intend to investigate the publications flagged by the *Nature* analysis. A spokesperson for Taylor & Francis said that some of the publications flagged were already under investigation by its ethics and integrity team.

When it comes to hallucinated references, "There have been cases where authors have been able to clearly document where issues have occurred in the process of producing a manuscript, for example using a translation tool, and demonstrate that the rest of the paper can be relied upon, in which case the paper will be corrected," says Chris Graf,

Springer Nature's research-integrity director. But, more often, these references reflect broader problems with the content, he says.

Shockman says that the number of potentially problematic citations flagged by Veracity is an order of magnitude greater when it is used in pilot programmes to screen submissions on behalf of publishers than when it analyses publications. This suggests that publishers are catching a large proportion of such citations before they can make it into the literature.

Nature's collaboration with Grounded AI also highlighted, as many experts have noted, that the detection of invalid citations with automated tools is not error-free. One of the challenges is that journals have various ways of formatting references, and AI tools might fail to recognize references because of how they are styled. These types of problem showed up among citations that manual checks determined to be genuine despite having been flagged by Grounded AI.

Another issue, says Weber-Boer, is that large-scale bibliometric databases might not index references that can't be verified, meaning their metadata might not match what appears on the publishers' websites. Some references do not contain their corresponding DOI, which makes it hard for automated tools to identify the cited paper, adds Weber-Boer. "We're starting to get a handle on the characteristics of this problem, which are a precursor to understanding the scale of it," she says.

The Grounded AI team members acknowledge that not all the references their tool flags will be true positives, but they say they are continuing to improve its performance. IOP Publishing, based in Bristol, UK, is now using Grounded AI's tool to screen submissions for problematic citations across all of its proprietary journals, says Kim Eggleton, head of peer

review and research integrity. “We know it’s a problem, we just don’t know how big the problem is,” she says.

Fake-citation fallout

Other start-up firms that are designing AI tools to catch fake references are also finding that they cannot yet eliminate manual checks. One such firm is GPTZero, based in New York City, which is working with the International Conference on Learning Representations (ICLR) on a tool for screening submissions.

GPTZero’s tool uses AI to search for the cited publications across the web as well as scholarly databases. Last year, the GPTZero team screened more than 700,000 citations in submissions to the ICLR 2026 and flagged 9,000 to be checked manually, says Alex Cui, the firm’s co-founder and chief technology officer. The flagged errors included titles that did not match those of any known publication, broken DOIs and citations attributed to authors with no connection to those works.

ICLR2026 programme chairs told *Nature* that around 1,000 manuscripts contained flagged citations. They rejected any that were found to contain hallucinated references, although they declined to say exactly how many.

Other efforts and approaches are also emerging. Michał Wójcik, who has just completed a PhD in neuroscience at the Free University of Berlin, says that he uses tools based on the programming language Python to screen references he samples from Crossref, looking for mismatches in their metadata. So far, he has identified more than 500 papers with unresolved or mismatched DOIs, which he checked manually and reported on PubPeer – a platform for post-publication review.

Another tool is CheckIfExist, an open-source platform that checks whether one or several references exist in scholarly databases. The tool, described in a preprint posted in January⁹, can help authors to avoid citing non-existent

publications. Cabanac has worked with the French national research agency CNRS in Paris to develop another tool, called bibCheck, which has been made freely available to CNRS scientists. It checks whether a reference corresponds to an existing or retracted work.

In Cabanac’s view, a paper with hallucinated citations should not be in the literature. It’s “like a plane flying with inexistent bolts”, he says. Once such a paper is identified, publishers must issue an expression of concern promptly and assess whether it needs to be corrected or retracted, he says.

In several cases over the past year, publishers have retracted papers and books that were found to contain hallucinated citations.

After Cabanac reported the *International Dental Journal*¹ paper with the invalid reference to PubPeer in January, it was corrected in March to cite his preprint. Brett Duane, a public-dental-health researcher at Trinity College Dublin in Ireland and a co-author of the paper, which was about using ChatGPT to estimate carbon emissions in dental practices, says that ChatGPT was used to “help identify potentially relevant references for use in the manuscript, which were all independently reviewed and verified”. The reference to Cabanac’s work was “a corrupted citation, originating from an LLM suggestion” that was inadvertently added during drafting, he says, and did not affect the paper’s conclusions.

Scholars are still debating whether and when hallucinated citations should be considered a form of research misconduct. Even major problems can happen unintentionally. For example, authors might use LLMs in a rush to format their reference list, unaware that AI has rephrased the citations, changed DOIs or added made-up references. Policies on AI use differ across journals. Most publishers require authors to disclose the use of AI, although what specific uses need to be disclosed varies. Policies also require human oversight of AI tools.

Failing to verify their output could represent a lack of scientific rigour, according to several researchers who spoke to *Nature*.

Even when the hallucinated citation doesn’t affect a paper’s findings and AI use is disclosed in accordance with journal policies, the journal must issue a public correction, Hosseini says.

In a paper published in March, he and David Resnik, a bioethicist at the National Institute of Environmental Health Sciences in Research Triangle Park, North Carolina, suggest that citations that point to non-existent work in reviews and bibliometric studies should be classified as a form of research misconduct if the references serve as data that directly support findings and conclusions¹⁰. In this case, the invalid citations amount to data fabrication, they argue.

In some cases, hallucinated citations might serve as a sign that the entire paper is fabricated, whether by an individual or a paper mill, a business that sells authorship slots. But the extent to which fabricated papers are contributing to the fake-citation problem is unclear.

Spokespeople for Sage and Taylor & Francis told *Nature* that when a hallucinated citation is found in a submission, they might reject the manuscript or ask for a revision if the errors don’t affect the conclusions and don’t suggest wider misconduct. A spokesperson for Wiley said that “minor issues may be raised to the author for clarification”. Springer Nature said it withdraws submissions that are found to contain hallucinated references.

Johnston says she is taking a hard line: when such a citation is found in a submission to *RIPE*, she says, the authors are prohibited from resubmitting the work to the journal.

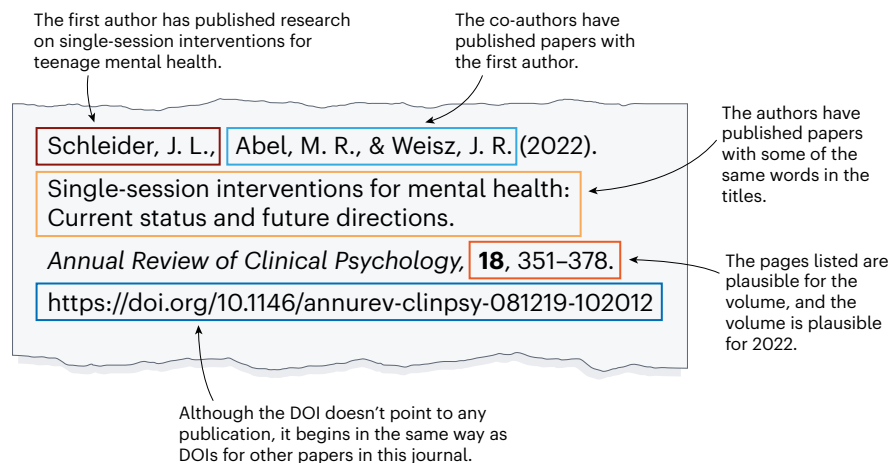
Hallucinated citations that make it into the academic literature can slow down and confuse other researchers’ efforts, and lead to false conclusions. Such errors can also create distrust in science, says Weber-Boer.

Hosseini agrees: “Every fake citation is a problem in the literature that someone would have to deal with.”

Miryam Naddaf is a science writer for *Nature* in London. **Elizabeth Quill** is a freelance editor in Washington DC.

HOW FAKES CAN LOOK REAL

This citation, generated by AI as part of a 2025 study, looks plausible even though it points to research that doesn’t exist.



1. Duane, B., Ashley, P. & Larkin, J. *Int. Dent. J.* **76**, 103979 (2026).
2. Cabanac, G., Labbé, C. & Magazinov, A. Preprint at arXiv <https://doi.org/10.48550/arXiv.2107.06751> (2021).
3. Sakai, Y., Kamigaito, H. & Watanabe, T. Preprint at arXiv <https://doi.org/10.48550/arXiv.2601.18724> (2026).
4. Bienz, A., Pearson, C. & Garcia de Gonzalo, S. Preprint at arXiv <https://doi.org/10.48550/arXiv.2602.05867> (2026).
5. Baethge, C. & Jergas, H. *Res. Integr. Peer Rev.* **10**, 13 (2025).
6. Cobb, C. L., Crumly, B., Montero-Zamora, P., Schwartz, S. J. & Martínez, C. R. *Jr Am. Psychol.* **79**, 299–311 (2024).
7. Linardon, J. et al. *JMIR Ment. Health* **12**, e80371 (2025).
8. Liang, W. et al. *Nature Hum. Behav.* **9**, 2599–2609 (2025).
9. Abbonato, D. Preprint at arXiv <https://doi.org/10.48550/arXiv.2602.15871> (2026).
10. Resnik, D. B. & Hosseini, M. *Account. Res.* <https://doi.org/10.1080/08989621.2026.2645390> (2026).