

AI and Research Integrity

Martin Martinez

Chancy.AI

Maxim Topaz is a Data Science professor at Columbia University. He studies health informatics. He is published in peer-reviewed journals. He knows how large language models work. In 2025, Topaz used an AI tool to help edit a paper he was writing. He was unaware the model had inserted a fabricated citation. It looked real. It was formatted correctly. It referenced a plausible author, a plausible journal, and a plausible year of publication. But it was fictional. Fortunately, his editorial staff caught it before the paper went to print.

"I was deeply embarrassed," Topaz told STAT News. "I checked for that, and it still almost happened to me."

That embarrassment became a research question. If a data scientist at Columbia could nearly publish a fabricated citation, what was happening across the millions of papers flowing through the biomedical literature every year?

The Topaz team audited 2.5 million papers and 97.1 million individual references indexed in PubMed Central. They used an automated citation verification system that cross-checked every reference against PubMed, Crossref, OpenAlex, and Google Scholar. They found 4,046 fabricated citations buried in 2,810 published papers.

In 2023, approximately one in 2,828 papers contained at least one fabricated reference. By 2025, that rate had climbed to one in 458 — a sixfold increase. In the first seven weeks of 2026 alone, the rate reached one in 277 papers. The rate of fake citations rose from roughly 4 per 10,000 papers in 2023 to approximately 57 per 10,000 by early 2026 — more than a twelvefold increase in three years.

Across the full audit, 246 individual papers contained three or more fabricated references, though the majority of the papers contained only one or two. Topaz believes the fabrications were unintentional, that researchers did not realize AI had slipped fake citations into their work. But the question remains: if references were fabricated, what else in those papers might have been created from pattern memory rather than factual evidence?

"One question that all of us have," Topaz told STAT News, "is: 'Is AI making science more efficient, helping us do better work, or even the same work, but faster, or is it just creating slop?'"

The fabrication problem in academic research was also revealed when GPTZero, a company that builds AI detection tools, analyzed 4,841 scientific papers accepted at one of the world's most prestigious AI conferences, Neural Information Processing Systems (NIPS) 2025. GPTZero found more than 100 fabricated citations in 53 published papers. Each paper had passed review by at least three experts. Under long-standing academic norms, even a single fabricated citation is grounds for rejection. Yet the fabricated papers had already been accepted, presented, and entered into the official scientific record. A similar analysis of papers under review at another leading AI conference, the International Conference on Learning Representations (ICLR) 2026, found 50 submissions with confirmed hallucinated citations. As with the NIPS 2025 fabrications, each paper at ICLR had already passed review by three to five human experts.

These are concerning findings: the world's leading AI researchers, the people who understand these systems better than anyone, are being fooled by the same hallucinations affecting undergraduates.

Nature published its own investigation in April 2026, concluding that tens of thousands of publications from 2025 alone might include invalid references generated by AI. As the journal's analysis notes, the peer review process was never designed to verify whether cited papers were based on fabrications, because until the advent of AI, there had never been a reason to question the integrity of the references.

Researchers at Springer, one of the world's largest academic publishers, documented what they called the "chain effect", whereby a single hallucinated citation enters the published record and contaminates future research. When a systematic review or clinical guideline cites a paper that contains fabricated references, the fabricated data proliferates. Meta-analyses build on studies and evidence that was never produced.

Springer's team identified four distinct types of academic hallucination: citation fabrication, factual distortion, logical inconsistency, and propagation errors. The last category is the most dangerous because it is the least visible. A fabricated citation in a single paper is discoverable. The same fabrication in subsequent meta-analyses becomes nearly invisible, even though it may be influencing treatment decisions.

In November 2025, researchers at Deakin University in Australia published a study in JMIR Mental Health that tested GPT-4o's ability to generate reviews on mental health. Of 176

citations produced, 19.9 percent were completely fabricated. They cited papers that simply do not exist. Among the 141 citations that referenced real papers, 45.4 percent contained serious errors. Combined, 56.2 percent of the citations were unusable for academic research.

Fabrication rates varied dramatically by topic. For well-researched conditions like major depressive disorder, only 6 percent of citations were fabricated. For specialized topics like body dysmorphic disorder, the rate jumped to 29 percent. The less training data available on a subject, the more likely AI may invent sources.

Recent analysis shows improvement in newer models but the problem remains. GPT-5 with web search integration reduces fabrication rates to approximately 7-8 percent — a meaningful improvement, but still a rate that produces seven or eight fake citations out of a hundred samples. Stanford research from 2026 found that combining retrieval-augmented generation reduced hallucination rates dramatically, but that technology has yet to be implemented in general practice.

In March 2026, a paper in Taylor & Francis's *Accountability in Research* journal reframed the conversation. The paper argues that AI-generated hallucinated citations are not merely errors or quality-control failures. Fabricated citations may constitute research misconduct under United States federal regulations.

In one paper published in the journal *Academic Ethics*, 19 of 29 citations were hallucinated. In a paper retracted by PLOS ONE, 18 of 76 were fabricated. In a piece published in *Digestive Diseases and Sciences*, 12 of 14 cited references were hallucinated. These were indexed, peer-reviewed publications in the scholarly record. ArXiv, the preprint server that hosts much of the world's cutting-edge research, is now banning authors for submitting hallucinated references.

Legal researcher Damien Charlotin maintains a database tracking court cases involving AI-generated fake citations. His database contained 120 cases in May 2025. By December 2025, it had grown to 660. By February 2026, it reached 946. By July 2026, it stood at 1,598. The database is currently adding four to five new cases every day. Meanwhile, judges are increasingly holding attorneys accountable for AI fabrications with penalties totaling \$145,000 in the first quarter of 2026.

There is no question that AI tools introduce fabricated citations. As a result, researchers who use AI are increasingly held responsible for their authenticity.

LINKED CITATIONS — In Order of Appearance

[1] STAT News. "Fraudulent Citations, Blamed on AI Hallucinations, Are Becoming More Common in Research Papers." May 7, 2026.

<https://www.statnews.com/2026/05/07/lancet-study-finds-steep-rise-fraudulent-citations-academic-papers/>

[2] The Lancet. "Fabricated Citations: An Audit Across 2.5 Million Biomedical Papers." May 8, 2026. [https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(26\)00603-3/fulltext](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(26)00603-3/fulltext)

[3] Retraction Watch. "One in 277 PubMed-Indexed Papers in 2026 Shows Fabricated References, Says Analysis." May 8, 2026.

<https://retractionwatch.com/2026/05/07/one-in-277-pubmed-indexed-papers-in-2026-shows-fabricated-references-says-analysis/>

[4] ResearchGate / Springer. "AI Hallucinations in Academic Writing: Implications for Research Integrity." May 30, 2026.

https://www.researchgate.net/publication/405472511_AI_hallucinations_in_academic_writing_implications_for_research_integrity

[5] Fortune. "AI Hallucinations Are Slipping Past Experts Into Papers and Books to Enter the Permanent Record." May 24, 2026. <https://fortune.com/2026/05/24/ai-hallucinations-scientific-research-authors-medical-journal-treatment/>

[6] Forbes (Nietzel). "AI Blamed for Rise in Fabricated Citations Found in Recent Research Papers." May 12, 2026.

<https://www.forbes.com/sites/michaelnietzel/2026/05/12/ai-blamed-for-rise-in-fabricated-citations-found-in-recent-research-papers/>

[7] Forbes (Lee). "Fabricated Citations Found in Over 2,800 Biomedical Journal Articles." May 30, 2026. <https://www.forbes.com/sites/brucelee/2026/05/30/ai-fabricated-citations-in-over-2800-biomedical-journal-articles/>

[8] Fortune / GPTZero. "NeurIPS Research Papers Contained 100+ AI-Hallucinated Citations." January 2026. <https://fortune.com/2026/01/21/neurips-ai-conferences-research-papers-hallucinations/>

[9] arXiv. "Compound Deception in Elite Peer Review: A Failure Mode Taxonomy of 100 Fabricated Citations at NeurIPS 2025." 2026. <https://arxiv.org/pdf/2602.05930>

[10] BetaKit. "Start-Up Investigation Reveals 50 Peer-Reviewed Papers Contained Hallucinated Citations." December 2025. <https://betakit.com/start-up-investigation-reveals-50-peer-reviewed-papers-contained-hallucinated-citations/>

[11] Nature. "Hallucinated Citations Are Polluting the Scientific Literature. What Can Be Done?" April 1, 2026. <https://www.nature.com/articles/d41586-026-00969-z>

[12] JMIR Mental Health (Linardon et al.). "Influence of Topic Familiarity and Prompt Specificity on Citation Fabrication." November 2025.
<https://mental.jmir.org/2025/1/e80371>

[13] INRA.AI. "AI Citation Hallucinations: Verify Fake References (2026)." June 2026.
<https://www.inra.ai/blog/citation-accuracy>

[14] Taylor & Francis / Accountability in Research. "Hallucinated Citations Produced by Generative AI May Constitute Research Misconduct." March 15, 2026.
<https://www.tandfonline.com/doi/full/10.1080/08989621.2026.2645390>

[15] Charlotin, Damien. "AI Hallucination Cases Database." Ongoing.
<https://www.damiencharlotin.com/hallucinations/>

[16] Futurism. "Academics in Meltdown Now That They're Responsible for AI Hallucinations." May 23, 2026. <https://futurism.com/artificial-intelligence/academics-meltdown-ai-hallucinations-research>