Confabulated references in the age of AI: contamination of the biomedical scientific literature

Luca Fiorillo[1,2]*

[1]Department of Medicine and Surgery, University of Enna "Kore", 94100 Enna, Italy
[2]Department of Dental Research Cell, Dr. D. Y. Patil Dental College & Hospital, Dr. D. Y. Patil Vidyapeeth (Deemed to be University), Pune 411018, Pimpri, India

***Correspondence:** Luca Fiorillo, Department of Medicine and Surgery, University of Enna "Kore", 94100 Enna, Italy.
lucafiorillo@live.it; luca.fiorillo@unikore.it

Abstract

Large language models (LLMs) like ChatGPT are increasingly used in drafting scientific papers. While they can improve clarity and efficiency, a troubling issue has emerged: the inclusion of fabricated references—nonexistent citations that can mislead, especially in biomedical research where evidence integrity is crucial. Studies indicate that 69% of references in ChatGPT's medical queries are false, and only 7% of AI-generated medical articles contain accurate references. These fake citations often mimic real authors and journals, making detection difficult. Such inaccuracies can compromise research integrity, skew citation metrics, and reduce trust in scientific literature. To address this, journals are adopting policies requiring disclosure of AI use and human verification of references. Nonetheless, detecting AI-related misinformation remains challenging, and many experts believe the problem is bigger than currently known. Going forward, authors should avoid relying solely on LLMs, and reviewers must scrutinize references carefully. The scientific community needs to balance AI's usefulness with rigorous oversight, ensuring that the pursuit of efficiency doesn't undermine credibility. Ultimately, safeguarding research from AI-generated misinformation will require combined efforts of transparency, vigilance, and adherence to ethical principles, preserving the integrity of biomedical science.

Keywords

large language models, scientific writing, systematic review, research integrity, fabricated references, ChatGPT

Introduction

The advent of artificial intelligence (AI)-assisted writing tools has sparked both excitement and debate in the academic world. Nowhere is this more evident than in the biomedical sciences, where clear and precise communication of data is paramount. Proponents of large language models (LLMs) argue that these tools

can improve the writing process, enhancing clarity, reducing human biases, and even catching analytical errors that authors might overlook [1, 2]. Indeed, when used responsibly, AI writing aids might help structure complex arguments and support greater precision in scientific manuscripts [3]. However, alongside these potential benefits, a serious concern has emerged regarding the rigor and integrity of AI-generated scientific texts. In particular, reports have surfaced of manuscripts written with the help of AI that contain fabricated references, sources listed in reference sections that do not exist in any database or journal [4].

This phenomenon is not merely theoretical. By 2023, numerous published papers across fields showed signs of undisclosed ChatGPT use, some going viral for their flaws [5]. In the medical literature, the problem is especially pernicious. Scientific publications are built on trustworthy citations; each reference should direct readers to prior evidence that supports the new work. If a fictitious reference breaks that chain of evidence, the credibility of the paper collapses. Worse, if other scientists base new research or clinical decisions on false data from a fabricated reference, the damage multiplies. This contamination of the knowledge base can be challenging to detect and undo. The following sections outline what is currently known about AI-related reference fabrication, provide examples of its occurrence, and discuss the implications for the biomedical research community.

The rise of "hallucinated" references by LLMs

LLMs do not retrieve information in the same way a search engine does; instead, they generate text that statistically resembles the patterns in their training data. As a result, when asked to provide citations or bibliographies, an LLM like ChatGPT may produce references that look legitimate, complete with author names, article titles, journal names, and even volume and page numbers, but which are, in reality, invented. Such invented citations have been termed "hallucinated references," and their occurrence is well documented.

In a recent study published in *Mayo Clinic Proceedings: Digital Health*, Gravel et al. [2] evaluated ChatGPT's answers to medical questions and the references it provided. The findings were sobering: out of 59 references generated by ChatGPT, 41 (69%) were fabricated, despite appearing superficially plausible [1, 2]. Most of these fake citations had elements of truth mixed in; for example, the reference might list real researchers who work in the relevant field and a reputable journal, but the combination (authors, title, journal, year) was entirely made-up. It could not be found in any literature database. The authors noted that ChatGPT's responses were otherwise articulate. Still, the inclusion of such a high proportion of bogus references poses a serious risk if one were to trust the AI's output uncritically [1, 2, 6].

Another independent assessment, by Bhattacharyya et al. [7] in the journal *Cureus*, underscored the prevalence of this issue across multiple medical topics. In that study, ChatGPT-3.5 was prompted to generate short academic papers (with references) on various healthcare subjects. Among 115 total references produced, 47% were entirely fabricated, and another 46%, while linking to honest articles, contained some incorrect details (such as wrong year, volume, or page numbers). Alarmingly, only 7% of AI-generated references were entirely accurate in that experiment. The most common errors included incorrect PubMed ID numbers (in 93% of the AI-generated references) and mismatched publication details [7]. These studies confirm that hallucinated citations are not rare but rather a frequent byproduct of using LLMs for academic content generation.

Crucially, the fictitious references produced by AI are often highly deceptive. They typically mimic the format of genuine citations and usually incorporate the names of established researchers in the field, along with plausible article titles. For instance, Gravel et al. [2] observed that many fabricated references used "names of authors with previous relevant publications" and credible-sounding titles in reputable journals. An unwary reader (or even a peer reviewer) might not immediately suspect these references are fake without actively checking them in a database. *Retraction Watch*, a forum that tracks scientific integrity issues, has noted that such nonexistent or error-laden citations have become a signature of LLM-generated text [5, 8]. In other words, a reference list filled with impressive-looking but untraceable citations is a red

flag that an AI may have drafted a manuscript. The underlying reason is apparent: rather than searching and retrieving actual bibliographic data, ChatGPT essentially guesses what a suitable reference might look like, based on patterns in its training data. This limitation of LLMs leads to what one might call "auto-plagiarism of non-existent sources", the model isn't stealing existing text, but creating scholarly facsimiles that have no grounding in reality.

Operational taxonomy of reference fabrication

To improve analytical precision, AI-related reference fabrication can be operationally divided into three primary types:

- Type I—Fully fabricated references: Non-existent articles or books invented by the LLM, often with plausible authors and journal titles.
- Type II—Authentic references with corrupted metadata: Real publications whose bibliographic fields (year, volume, DOI, or page numbers) are distorted or mismatched.
- Type III—Chimeric or misattributed references: Citations combining elements of different authentic records (e.g., real authors with incorrect article titles or journals).

This taxonomy enables a standardized approach to detection and categorization across editorial, bibliometric, and informatics settings.

Prevalence and error patterns

Empirical assessments across biomedical and scientific corpora demonstrate alarming error frequencies. Gravel et al. [2] observed that ChatGPT generated fabricated citations in 48% of medical responses, while Walters and Wilder [9] documented approximately 30% fabricated and 20% erroneous references in automatically produced bibliographies. Bhattacharyya et al. [7] similarly reported high rates of unverifiable citations across biomedical queries. These findings confirm that AI confabulation of references remains a non-trivial and reproducible pattern, highlighting the urgent need for AI-output verification workflows.

Layered impacts of fabricated references

The implications of AI-induced reference fabrication are multifaceted:

1. Epistemic harms: Misleading researchers through false attribution of data or methods, undermining the reliability of background sections, and evidence mapping.
2. Synthesis harms: Contamination of systematic reviews and meta-analyses when spurious sources are unknowingly included, distorting pooled estimates and conclusions.
3. Clinical and policy harms: Risk of flawed guideline formulation or patient-care recommendations derived from compromised secondary evidence.
4. Publishing ecosystem harms: Increased workload for reviewers and editors, reputational damage to journals, and erosion of public trust in scientific integrity.

Documented cases of fabricated references

What was initially a hypothetical concern has now manifested in concrete cases across the scientific literature. Over the last two years, several published works have been corrected or retracted after it was discovered that they contained references to non-existent studies. These incidents span different disciplines (including biomedical fields) and illustrate how pervasive the problem has become in the wake of rapid AI adoption [9].

One early cautionary example emerged in the field of public health and environmental science. In March 2024, the journal *Environment and Planning E: Nature and Space* published an article on wildlife disease policy that, upon post-publication review, was found to contain "several non-existent citations," including one referring to a paper in 2026 [10]. The impossibility of referencing 2026 tipped readers off

that something was amiss. A wildlife ecologist who examined the paper noted that numerous citations to reputable sources were "incorrectly cited, abjectly false, or obviously manufactured" within the reference list [10]. This prompted an investigation by the journal's editors. The inquiry revealed that the lead author, unbeknownst to the other co-authors, had used ChatGPT to help format and "update" the references before submission [10]. In doing so, the author apparently trusted the AI to generate or correct citations, but instead, it introduced errors and fake entries. The original submission of the manuscript included accurate references (drawn from the author's Master's thesis), but the AI-altered version swapped in bogus sources that initially passed editorial checks [10]. The journal treated this issue seriously and announced it would take corrective actions. Fortunately, in this case, the authors provided the correct reference list once the problem was identified, and the paper could potentially be amended rather than fully retracted. Nonetheless, for a time, the paper was in the literature with its defective references, meaning anyone who read it shortly after publication could have been misled about the supporting sources of its claims.

Another case involved an analysis of communication and media studies that was suspected to be heavily AI-generated. In 2023, a professor reviewing a submission on community radio noted something uncanny: the writing style felt like AI produced it, and several references were questionable [11]. Among them was a citation of a 2017 article purportedly written by the reviewer herself, yet she knew she had published nothing on that topic after 2012. Indeed, the cited journal had no record of the 2017 paper. This indicated that the reference was entirely fabricated, presumably inserted by an AI tool used by the authors. The paper, initially rejected after peer review, resurfaced in another journal with minimal changes months later, suggesting that the authors attempted to publish the AI-written content elsewhere. Once again, the sharp-eyed reviewer (alerted by an automated system to the new publication) intervened, and the journal had to withdraw the article pending investigation. The authors in this incident defended themselves by citing the use of "bibliometric tools" and unstable indexing in repositories, but the glaring presence of a nonexistent citation attributed to the reviewer undermined their explanation [11]. This saga highlights how AI-generated errors can slip into the literature through less vigilant outlets, and how fabricated references can even implicate real researchers by name, creating confusion and reputational concerns.

Perhaps the most high-profile example to date is the case of an AI-generated book manuscript. In April 2025, Springer Nature published an e-book titled "*Mastering Machine Learning: From Basics to Advanced*". The book was marketed as an introductory text on machine learning, priced at a premium. However, readers soon noticed irregularities in the reference list. Upon closer examination by *Retraction Watch* and others, it was found that out of 46 references in the book, two-thirds either did not exist or contained significant inaccuracies [5]. Multiple researchers whose names appeared in the citations confirmed that the works attributed to them were entirely fictitious. For example, one cited paper was listed as appearing in a well-known IEEE journal. Still, the researcher, who is named as the author, clarified that, although he had a related unpublished preprint, he had never published in that journal. Another citation in the book referenced a section of a deep learning textbook that, in reality, had no such content on the cited pages. These fabrications were classic LLM confabulations, constructed from snippets of factual information assembled into a fabricated reference. The deception went unnoticed through the publication process, the book did not disclose any AI involvement in its creation (despite containing a section on ChatGPT ethics). When confronted with the allegations, the book's author did not admit to using AI. Still, he remarked on the difficulty of reliably determining AI-generated content and noted the challenge would only grow as AI becomes more sophisticated. Springer Nature, in response, stated that they have policies requiring authors to declare AI assistance and emphasize human oversight on all submissions [5]. Because these policies were evidently not followed or enforced in this case, Springer Nature retracted the book in July 2025, citing the discovery that it "referenced works that don't exist" in its citations [12]. This is one of the first significant book retractions due to AI-generated content. It underscores that even well-resourced publishers are vulnerable to AI deception if proper checks are not in place. Moreover, it serves as a stark warning that the scholarly record can be compromised on a large scale; a single book with dozens of fake references could misdirect countless readers or researchers until the fraud is uncovered [13].

Notably, not only text authors but also some peer reviewers have been implicated in using AI, potentially exacerbating the problem. Comments on the *Retraction Watch* coverage include anecdotes of reviewers submitting reports rife with AI-generated prose and erroneous citations [11]. If a reviewer were to "correct" an author's references using ChatGPT, for instance, they might unintentionally replace real citations with fake ones. This scenario, although anecdotal, highlights the multifaceted risk: AI misuse can enter the publication pipeline at multiple points (author, reviewer, or even editor), thereby increasing the likelihood that fallacious references will slip through.

Implications for biomedical research and evidence synthesis

The biomedical field is built on a hierarchy of evidence, with systematic reviews and meta-analyses occupying a prominent position in shaping practice and policy. These comprehensive reviews depend on the assumption that all included studies are real, and their data are accurately reported. The emergence of AI-generated references challenges this assumption and raises several concerning implications.

First, consider the integrity of a systematic review article itself. A systematic review typically involves identifying and screening hundreds or thousands of publications to include a curated set of relevant studies. If an author uses an AI tool to assist with this process, there is a risk that the tool may generate references to studies that appear relevant but do not actually exist. An unscrupulous or naive author might then include these phantom studies in the review's analysis. The review could report, for example, "10 randomized trials with a combined 2,000 patients," supporting a particular therapy, a convincing sample size, when in fact some of those trials were not actual clinical trials. The robust-sounding evidence might convince readers of the review (and even peer reviewers). The danger is that downstream researchers could cite this systematic review or even include it in a meta-analysis, unwittingly proliferating the falsehood. In essence, one AI-seeded lie in a reference list can grow into a forest of misinformation as papers draw upon each other. This cascading effect is how the "fallacy of information" mentioned earlier can take hold, creating a veneer of evidence where none exists.

We are likely already seeing early signs of such contamination. *Retraction Watch* maintains a list of papers with suspected undisclosed AI-generated content, and many are in biomedical domains (medicine, biology, psychology) [11]. Some meta-analyses have had to be corrected or retracted because they included data from papers that were later retracted for misconduct. If those retractions were due to fake references or AI confabulations, the meta-analytic conclusions would have been distorted. Even legitimate meta-analyses may be skewed if they unknowingly incorporate results from systematic reviews that are tainted. The full extent of this problem remains hidden due to the sheer difficulty of manually verifying every reference in every paper.

Another important implication is the erosion of trust. Science operates on a mix of trust and verification: we trust that authors cite real, relevant studies, and we verify key findings when necessary. If readers and editors must now approach every reference list with skepticism, the efficiency of scientific communication plummets. A commentary [14] in *The Scientist* noted that ChatGPT (as of 2023) failed to flag known retracted or debunked papers and even rated them highly credible, illustrating that AI is oblivious to the reliability of sources. Now, if AI is also fabricating sources wholesale, the trust deficit widens further. Researchers may start doubting even genuine references, especially if an article's prose gives any hint of AI-style generation. This could lead to more labor-intensive peer review and editorial processes, as humans must double-check what was once taken for granted.

From an ethical standpoint, introducing fabricated references into a paper constitutes academic misconduct. It violates principles of honesty and accuracy in scholarship. Some might argue that if done unintentionally via AI, it is an "honest mistake." However, the scale and foreseeability of this issue mean that researchers have a responsibility to actively guard against it. Using an AI tool without verifying its outputs is not an excuse—just as claiming "the software made a mistake" would not absolve a scientist who reports incorrect data analysis without validation. The community is beginning to discuss whether failure to verify AI outputs that lead to the propagation of false information should itself be considered negligence

or misconduct. Journals are certainly treating it seriously: the above cases demonstrate that publishers will retract or correct literature once fake references come to light, to protect the integrity of the scientific record [14].

Moreover, fabricated citations have implications that extend beyond the content of a single paper; they also impact bibliometrics and the broader scholarly ecosystem. As Bret Collier (the ecologist who raised the alarm on the wildlife paper) pointed out, manufactured citations can impact journal metrics and author indices [10]. In the short term, a fake reference might cite a non-existent article, doing no direct harm to metrics. But sometimes these AI-made references partially cite honest articles (e.g., wrong year or journal but correct author). In such cases, a real researcher's work could be misattributed and not adequately credited, skewing citation counts. Alternatively, the fake references might cite a real journal and give it a fabricated article, potentially confusing citation trackers and databases. If such references slip into databases like Google Scholar or Scopus before being caught, they could artificially inflate citation counts or create phantom entries that clutter search results.

Ultimately, the phenomenon necessitates a reevaluation of how we train and mentor researchers. The ease of generating text (including references) with AI might tempt students or less experienced authors to bypass important scholarly habits, such as thoroughly reading the literature. The adage "never cite a paper you haven't read" is being updated in the AI era to "never trust a citation you didn't verify." There is now an urgent need to instill in researchers the discipline of verifying every reference, whether it comes from an AI tool or even from another published paper. As one commentator quipped in response to these incidents: "*Call me old-fashioned, but the best way to cite papers? Read them first, then you know they are not a figment of either your imagination or that of a confabulating LLM.*" [11].

Preventive measures and call to action

Confronting this challenge will require both technological and cultural changes in the field of scientific publishing. On the technological side, tools are being developed to detect AI-generated text and even flag potentially fake references. For example, some reference manager software can be augmented to cross-check citations in a manuscript against databases automatically. Journals are beginning to use plagiarism-detection and AI-detection services on submissions, although these are not foolproof and can be bypassed or yield false positives. A proposed system would require that all references in submitted manuscripts be accompanied by DOIs or database IDs that can be automatically verified; a missing or invalid identifier would immediately raise a red flag for editors to investigate. Initiatives like the *xFakeSci* project are exploring machine learning algorithms to identify AI-generated scientific writing by subtle anomalies in text and referenc. In parallel, publishers like Elsevier and Springer Nature have introduced policies for AI: authors must declare AI assistance and are responsible for the content (including the accuracy of references) when AI is used [7, 12]. Such policies put the onus on authors to be transparent and careful [15].

However, no policy or tool can replace the fundamental role of human oversight and integrity. Authors who choose to use AI in writing must do so responsibly, as *assistive* technology rather than a source of truth. This means rigorously checking every reference ChatGPT suggests, just as one would double-check a reference contributed by another co-author. Researchers need to be educated about the specific pitfalls of LLMs. The fact that ChatGPT can output an entirely fictional and very compelling-sounding reference is not common knowledge among all academics; therefore, raising awareness is critical. Conferences, editorials, and university guidelines should now include warnings about AI confabulations. The present article, in fact, serves as a commentary to raise such awareness in the biomedical community [13].

Peer reviewers and editors, the gatekeepers of scientific quality, must also adapt. It may no longer be sufficient to skim an article and assume the references are legitimate if the prose is acceptable. Some journals have started implementing extra reference checks. For instance, as one online discussion noted, certain publishers link cited references to databases during typesetting, which can catch references that don't match anything in CrossRef or PubMed [10]. This is a good practice that should become standard. If a

citation cannot be resolved to a real document, it should be queried with the authors. Additionally, journals could explicitly ask reviewers to assess the reference section for any irregularities. While this goes beyond traditional peer review expectations, it is a necessary evolution in response to new threats to quality [16].

On a systemic level, when fake reference papers are discovered, journals should issue swift corrections or retractions to prevent misinformation from spreading. *Retraction Watch*'s database is littered with examples of how slow or reluctant action can allow flawed science to continue influencing new work. Even Stanford University's list of the top 2% of researchers in the world begins to include the retracted articles for each author in a clear and public way [17]. In the cases described earlier, interventions by alert individuals (reviewers, readers) were essential. We cannot rely solely on serendipity or whistleblowers; formal mechanisms must be in place. Journals might consider post-publication audits for papers in which AI use is suspected, or even random spot-checks of references in a sample of publications.

The scientific community may also benefit from a more open dialogue about the acceptable use of AI in publishing. AI is here to stay, and banning it outright is neither feasible nor productive. Instead, as Fiorillo (2024) [1] argues, we should "embrace AI tools, rather than demonize them," but with a *commitment to integrity and ethical guidelines*. This balanced approach means encouraging innovation (like using AI to improve writing clarity or to generate hypotheses), while unequivocally condemning and penalizing careless or fraudulent use (such as dumping AI text and fake citations into a manuscript). It also means providing researchers with the training to use AI effectively. For example, using ChatGPT to proofread grammar is fundamentally different from using it to generate literature reviews or reference lists; the former carries little risk. At the same time, the latter can inject falsehoods if unchecked [18, 19].

As a community, scientists must reinforce a culture where intellectual honesty is paramount. The allure of easy writing should not outweigh the duty we have to truth and accuracy. Junior researchers in particular should be mentored in these values. Just as statistical falsification or image manipulation is not tolerated, we should view the introduction of invented references as a grave misstep, whether born of malice or ignorance. Only with clear expectations and education can we prevent a generation of researchers from normalizing the practice of "write now, verify never." A promising example of author-driven tools that may contribute to preserving research integrity is the recently introduced Fi-index, developed to limit self-citation practices and enhance bibliometric reliability. Although primarily conceived to address citation bias, a semi-automated method that requires active verification and effort from authors could also represent a valuable strategy to ensure the validity of cited references and mitigate the spread of fabricated sources [20].

Finally, readers of scientific literature (including clinicians who consume evidence to guide practice) should stay vigilant. If something in a paper seems too convenient or too neatly in support of the authors' narrative, especially a reference that is hard to find or oddly formatted, it might warrant a quick check. The democratization of knowledge means that the wider community, post-publication, can catch many such errors. Authors and journals should welcome this approach and treat readers who flag concerns as allies in upholding scientific quality, rather than adversaries [21].

Publisher and governance responses

Major governance bodies and publishers have introduced preliminary safeguards. The International Committee of Medical Journal Editors (ICMJE) (2024) explicitly mandates that AI-assisted technologies must be disclosed and cannot be listed as authors. COPE (2024) emphasizes transparent reporting and editorial responsibility in AI-assisted writing. Elsevier and Nature Portfolio now require formal disclosure of AI tools used in manuscript generation and recommend human verification of all bibliographic elements. While some publishers have implemented automated reference-checking algorithms, there are currently no publicly available, aggregated statistics on manuscript rejections attributable to AI-generated, fabricated references. This lack of systematic reporting remains a significant evidence gap in scholarly publishing (Table 1).

Table 1. Preventive and corrective strategies should be structured across complementary domains and stakeholder levels.

Stakeholder	Technical controls	Institutional/Editorial frameworks	Ethical education and training	Cross-institutional collaboration
Authors	Validate all DOIs and PubMed IDs; manually verify citations.	Include AI disclosure statements in manuscripts.	Attend AI integrity workshops.	Participate in shared bibliometric initiatives.
Reviewers	Cross-check random subset of references during peer review.	Report unverifiable citations to editors.	Encourage authors' transparency.	Support creation of shared blacklists of fabricated entries.
Editors and journals	Deploy automated bibliographic validation tools.	Adopt explicit AI-use policies; enforce author accountability.	Offer reviewer training on AI detection.	Collaborate with COPE and Crossref on metadata audits.
Publishers	Integrate LLM-output screening pipelines.	Require structured author contribution and verification statements.	Provide editorial-board guidance documents.	Coordinate global data-integrity task forces.
Readers and institutions	Verify doubtful citations through digital identifiers.	Promote awareness campaigns about AI confabulations.	Foster critical appraisal skills.	Encourage open post-publication peer review.

LLM: large language model.

Integrated preventive framework

This multi-tiered approach aligns technical verification with ethical accountability, ensuring that AI tools remain aids rather than sources of epistemic risk.

Conclusion

AI language models, such as ChatGPT, offer exciting possibilities for accelerating literature searches, drafting manuscripts, and translating scientific findings for broader audiences. Yet, as this commentary has detailed, they also introduce new perils into the publication ecosystem, chief among them, the generation of fictional references that can subvert the very foundation of evidence-based discourse. The recent surge in papers with AI-written sections has already led to a corresponding uptick in false citations polluting the scholarly record. In biomedical research, where lives and public health policies can hinge on published data, the stakes of this misinformation are exceptionally high.

The scientific community must address this issue proactively. This will involve technological solutions (improved reference verification and AI-detection tools) and, more importantly, cultural shifts in how we write and review manuscripts. All stakeholders, authors, reviewers, editors, publishers, and readers, must exercise increased diligence. By reasserting the basics (read what you cite and verify what you write) and combining them with new checks tailored to the AI era, we can prevent the worst outcomes. The volume of retractions required to rectify existing AI-induced errors may be substantial. Still, it is a necessary step to purge the literature of fraudulent information and send a message that quality control is catching up with technology.

In conclusion, AI can be a valuable ally in scientific writing if used wisely, but it cannot be trusted blindly. The occurrence of hallucinated references is a glaring reminder that human expertise, our capacity for critical thinking, skepticism, and verification, remains indispensable. Preserving trust in science will depend on our ability to integrate powerful tools like LLMs without compromising the core principles of scholarship. Through collective effort and adherence to integrity, the research community can harness the benefits of AI while mitigating its tendency to generate persuasive falsehoods. As we navigate this new era, let us reaffirm that factual accuracy is non-negotiable. The credibility of biomedical science rests on it.

Abbreviations

AI: artificial intelligence

LLMs: large language models

Declarations

Author contributions

LF: Conceptualization, Investigation, Writing—original draft, Writing—review & editing. The author confirms sole responsibility for all aspects of this paper, including the accuracy of the content and the integrity of the references cited. The author read and approved the submitted version.

Conflicts of interest

The author is the Editorial Board Member and Guest Editor of Exploration of Medicine, but this manuscript is handled separately by other editors, and the author is not involved in the decision-making or the review process.

Ethical approval

Not applicable.

Consent to participate

Not applicable.

Consent to publication

Not applicable.

Availability of data and materials

No external unpublished data, all data is published within the manuscript.

Funding

Not applicable.

Copyright

© The Author(s) 2026.

Publisher's note

Open Exploration maintains a neutral stance on jurisdictional claims in published institutional affiliations and maps. All opinions expressed in this article are the personal views of the author(s) and do not represent the stance of the editorial team or the publisher.

References

1. Fiorillo L. Confronting the demonization of AI writing: Reevaluating its role in upholding scientific integrity. Oral Oncol Rep. 2024;12:100685. [DOI]
2. Gravel J, D'Amours-Gravel M, Osmanlliu E. Learning to Fake It: Limited Responses and Fabricated References Provided by ChatGPT for Medical Questions. Mayo Clin Proc Digit Health. 2023;1:226–34. [DOI] [PubMed] [PMC]
3. Bai Y, Kosonocky CW, Wang JZ. How our authors are using AI tools in manuscript writing. Patterns (N Y). 2024;5:101075. [DOI] [PubMed] [PMC]
4. Mac Oscar M, Boland M, Cadogan C. An assessment of generative artificial intelligence in responding to clinical queries on tapering antidepressants. Res Social Adm Pharm. 2025;21:924–30. [DOI] [PubMed]
5. Madhavan G. RETRACTED BOOK: Mastering Machine Learning: From Basics to Advanced. In: Transactions on Computer Systems and Networks. Singapore: Springer Nature; 2025. [DOI]
6. Fiorillo L. Fi-Index: A New Method to Evaluate Authors Hirsch-Index Reliability. Publ Res Q. 2022;38: 465–74. [DOI]

7. Bhattacharyya M, Miller VM, Bhattacharyya D, Miller LE. High Rates of Fabricated and Inaccurate References in ChatGPT-Generated Medical Content. Cureus. 2023;15:e39238. [DOI] [PubMed] [PMC]

8. Glynn A. Guarding against artificial intelligence--hallucinated citations: The case for full-text reference deposit. arXiv 2503.19848 [Preprint]. 2025 [cited 2025 Nov 25]. Available from: https://arxiv.org/abs/2503.19848

9. Walters WH, Wilder EI. Fabrication and errors in the bibliographic citations generated by ChatGPT. Sci Rep. 2023;13:14045. [DOI] [PubMed] [PMC]

10. Journal taking 'corrective actions' after learning author used ChatGPT to update references [Internet]. Retraction Watch; [cited 2025 Sep 21]. Available from: https://retractionwatch.com/2024/05/20/journal-taking-corrective-actions-after-learning-author-used-chatgpt-to-update-references/

11. Paper rejected for AI, fake references published elsewhere with hardly anything changed [Internet]. Retraction Watch; [cited 2025 Sep 21]. Available from: https://retractionwatch.com/2025/06/12/paper-rejected-for-ai-fake-references-published-elsewhere-with-hardly-anything-changed/

12. Springer Nature retracts machine learning book after citations 'reference works that don't exist' [Internet]. The Bookseller; [cited 2025 Sep 21]. Available from: https://www.thebookseller.com/news/springer-nature-retracts-machine-learning-book-after-citations-reference-works-that-dont-exist

13. Kalai AT, Nachum O, Vempala SS, Zhang E. Why Language Models Hallucinate. arXiv 2509.04664 [Preprint]. 2025 [cited 2025 Nov 25]. Available from: https://arxiv.org/abs/2509.04664

14. ChatGPT Fails to Flag Retracted and Problematic Articles [Internet]. The Scientist; [cited 2025 Sep 21]. Available from: https://www.the-scientist.com/chatgpt-fails-to-flag-retracted-and-problematic-articles-73448

15. Wallace MB, Siersema PD. Ethics in publication. Gastrointest Endosc. 2015;82:439–42. [DOI] [PubMed]

16. Singhal S, Kalra BS. Publication ethics: Role and responsibility of authors. Indian J Gastroenterol. 2021;40:65–71. [DOI] [PubMed] [PMC]

17. Ioannidis JPA, Boyack KW, Baas J. Updated science-wide author databases of standardized citation indicators. PLoS Biol. 2020;18:e3000918. [DOI] [PubMed] [PMC]

18. Yoo J. Defining the Boundaries of AI Use in Scientific Writing: A Comparative Review of Editorial Policies. J Korean Med Sci. 2025;40:e187. [DOI] [PubMed] [PMC]

19. Chetwynd E. Ethical Use of Artificial Intelligence for Scientific Writing: Current Trends. J Hum Lact. 2024;40:211–5. [DOI] [PubMed] [PMC]

20. Fiorillo L, Cicciù M. The Use of Fi-Index Tool to Assess Per-manuscript Self-citations. Publ Res Q. 2022;38:684–92. [DOI]

21. Smith R. US calls for research into research integrity. BMJ. 2000;321:1369B. [PubMed] [PMC]