Biointerface Research in Applied Chemistry
Open-Access Journal (ISSN: 2069-5837)

Editorial
Volume 16, Issue 4, 2026, 101
https://doi.org/10.33263/BRIAC164.101

Indexed, Ranked, Accused: Why Bibliometric Status Is Not a Certificate of Integrity in the Age of AI-Hallucinated Citations

Alexandru Mihai Grumezescu [1,*]

[1] Department of Science and Engineering of Oxide Materials and Nanomaterials, Faculty of Chemical Engineering and Biotechnologies, National University of Science and Technology POLITEHNICA Bucharest, 1-7 Gheorghe Polizu Street, 011061 Bucharest, Romania; agrumezescu@upb.ro (A.M.G.);
[2] Research Institute of the University of Bucharest, University of Bucharest, 90 Panduri Street, 050663 Bucharest, Romania
* Correspondence: agrumezescu@upb.ro (A.M.G.);

Received: 30.05.2026; Accepted: 10.06.2026; Published: 15.06.2026

Preamble

In a series of recent editorials, we have reflected on the evolving architecture of scholarly evaluation, the asymmetries that may emerge when standards are applied unevenly across journals, and the distinction between rigor as a declared principle and consistency as its operational condition.

In “Rigor or Symmetry? Reflections on Fifteen Years of Diamond Open Access” (doi:10.33263/BRIAC161.001), the focus was the asymmetry between entry and presence: journals applying or reapplying for evaluation may face increasingly granular technical thresholds, while established journals often continue to benefit from historical inclusion. In “Consistency or Contingency? Reflections on Uncertainty in Editorial Triage” (doi:10.33263/LIANBS151.001), the issue was the variability of emphasis across comparable evaluations. In “Consistency or Context? Reflections on Indexing, Evaluation, and Temporal Validity” (doi:10.33263/BRIAC162.070), the question was whether the same scholarly archive can occupy different evaluative positions depending on timing, database architecture, or procedural context. In “Broken Links, Broken Symmetry? Reflections on Technical Formalism and Evaluative Reciprocity” (doi:10.33263/BRIAC163.071), attention turned to the proportional interpretation of technical imperfections. In “Applied Chemistry or Applied Everything? Reflections on Scope, Citation, and Evaluative Coherence” (doi:10.33263/BRIAC163.100), the focus shifted to the relationship between category mapping, disciplinary boundaries, and evaluative coherence.

The present editorial extends that sequence toward a different but increasingly urgent question: citation integrity in indexed journals.

More specifically, it asks what happens when fabricated, unverifiable, or AI-hallucinated references appear in scholarly articles that have passed through editorial workflows and have been published in journals included in major indexing systems such as Web of Science.

This question is not intended as an accusation against any particular journal, publisher, evaluator, or database. Nor is it an argument against the use of artificial intelligence in scholarly writing. Generative tools are already part of contemporary academic workflows, and their presence will only increase.

The issue is not whether artificial intelligence exists in the writing process.

The issue is whether the scholarly record can still distinguish between real sources and plausible imitations of sources.

This distinction does not concern references alone. AI-hallucinated citations are the most visible and easily verifiable symptom of a broader problem: the emergence of scholarly-looking text that may imitate the language of research without being anchored in verifiable scholarly substance. A fabricated reference is one form of hallucination. An unsupported claim, a simulated literature consensus, an invented methodological justification, or a misrepresented source may represent another. The reference list is where the defect becomes easiest to expose, but the vulnerability may begin earlier, inside the argument itself.

References as Scholarly Infrastructure

A scholarly article does not exist in isolation. It is connected to previous work through references. These references are not decorative elements, nor merely formal requirements added at the end of a manuscript. They are the pathways through which claims can be verified, intellectual debts can be traced, methods can be compared, and scientific arguments can be situated within an existing body of knowledge.

A reference tells the reader: this claim, method, observation, or interpretation is connected to something that exists outside the present text.

For this reason, citation integrity is not a minor technical matter. It is part of the epistemic infrastructure of science. The trustworthiness of a scholarly article depends not only on the internal coherence of its argument, but also on the external reality of the works to which it points.

When references are fabricated, this pathway collapses.

The problem is therefore not merely one of citation style, editorial formatting, or bibliographic polish. It is a question of whether the scholarly record still points to real scholarly objects.

A citation error may misdescribe the scholarly record.

A hallucinated reference fabricates it.

The Emergence of a Scalable Problem

Fabricated references are not new. Long before the emergence of large language models, scholarly publishing had already encountered cases of articles containing incoherent content, invented sources, manipulated citations, or bibliographic entries that should have been detected through basic editorial scrutiny.

What has changed is scalability and plausibility.

Large language models are capable of generating references that appear structurally credible while being factually nonexistent. A hallucinated reference may contain real authors associated with false titles, real journals with nonexistent articles, plausible volumes and page ranges, syntactically valid but non-resolving DOI numbers, or article titles that sound entirely appropriate to the field but do not correspond to any published work.

The danger lies precisely in this plausibility.

The fake reference does not necessarily look absurd. It may look normal. It may fit the argument. It may be formatted correctly. It may even contain enough real fragments to evade superficial inspection.

A fabricated citation generated by AI is therefore not simply a visible error. It may be an invisible absence. The article remains fluent. The paragraph appears supported. The reference list appears complete.

Only verification reveals that the support is not there.

The scale of the problem is no longer anecdotal. Recent large-scale analyses have moved the discussion from isolated examples to measurable contamination. One audit of more than one hundred million references across millions of papers and preprints estimated nearly one hundred and fifty thousand hallucinated citations in 2025 alone. A separate biomedical analysis identified thousands of fabricated references across thousands of papers, with the estimated prevalence rising sharply between 2023, 2025, and the first weeks of 2026. These figures do not describe occasional embarrassment. They describe a verification problem entering the scholarly record at scale [1-3].

This is why fabricated references are so revealing. They are not merely errors at the margin of a manuscript; they may indicate that the surrounding text has also been produced through a process of simulation rather than verification. If a paragraph is supported by a nonexistent source, the problem is not only that the source is fake. The problem is that the claim itself may have entered the manuscript without passing through contact with the scholarly record. In such cases, hallucinated citations become diagnostic markers of a deeper epistemic failure.

This raises a difficult question.

If the scientific community now recognizes AI-hallucinated references as a foreseeable risk, can their detection remain optional?

The *Metalurgia International* Precedent

The problem of fabricated references did not begin with artificial intelligence.

A useful historical precedent is the 2013 case of *Metalurgia International*, which published the deliberately fabricated article “Evaluation of Transformative Hermeneutic Heuristics for Processing Random Data.” [4] (available in publicly accessible online copies, including Scribd). The article was not merely weak, eccentric, or poorly written. It was constructed as a parody of pseudo-scientific writing. Its title, abstract, terminology, figures, author presentation, and bibliography were intentionally implausible, yet the article passed through the journal’s publication process and appeared as a scholarly contribution.

The significance of the case lies not only in the absurdity of the article, but in the failure of the editorial system that allowed it to become part of the published record. The bibliography did not simply contain minor errors. It contained a fabricated or parodic bibliographic universe: implausible scholarly objects, suspicious author-source combinations, and references that should have raised immediate concern under even minimal editorial scrutiny.

At the time, this was not an AI problem.

It was a verification problem.

The case is important because *Metalurgia International* was not merely an invisible publication venue outside the evaluative ecosystem. It had bibliometric visibility and had previously been associated with indexed status.

The journal had an Impact Factor for 2012 and, according to publicly available records, ceased to be referenced in Web of Science in 2013; it was also reported as having been removed from SCI Expanded with the status “Dropped.” Whether this outcome reflected the single scandal, a broader accumulation of editorial concerns, or a wider reevaluation of journal

quality, the structural lesson remains clear: fabricated scholarship passing through publication workflows was treated as a serious sign of editorial failure.

In that case, the response of the indexing ecosystem was decisive: the journal lost its position within the Web of Science/JCR framework. More than a decade later, the question is whether the same logic will be applied to a more sophisticated version of the same defect. If fabricated bibliographic structures were followed by decisive editorial and indexing consequences when they appeared in visibly absurd form, should AI-generated fabricated references be treated differently merely because they are more plausible, better formatted, and less immediately detectable?

AI-hallucinated citations may look more polished than the absurd references of the *Metalurgia International* episode. They may use plausible journal titles, credible author names, realistic article titles, coherent volume and page numbers, and syntactically convincing DOI patterns. They may not look like parody. They may look like normal scholarship.

But cosmetic plausibility does not change the nature of the defect.

A fake reference remains fake whether it appears absurd or elegant. A fabricated bibliographic object remains fabricated whether it is produced manually, satirically, negligently, or by a language model. The difference is not epistemic. It is operational. What once appeared as an obvious failure of editorial scrutiny can now enter the literature in a subtler and more scalable form.

The comparison also changes when scale is considered. The *Metalurgia International* episode was memorable because it exposed a spectacular and visible editorial failure. AI-hallucinated references, by contrast, may appear as distributed micro-failures across journals, repositories, conferences, and disciplines. Their danger lies not only in individual fabrication, but in accumulation: thousands of small bibliographic absences becoming part of the searchable scholarly record [1,2].

This is why the comparison matters.

If a visibly fabricated article with visibly problematic references could be treated as evidence of serious editorial weakness in 2013, then AI-generated fabricated references should not be normalized in 2026 merely because they are better written. The surface has improved. The underlying failure has not changed.

The difference between the *Metalurgia International* episode and the contemporary AI-hallucination problem is not that one involved fabrication and the other does not. Both involve the entry of fabricated scholarly structures into the publication process. The difference is aesthetic and operational. In 2013, the fabrication was visible because it was parodic, excessive, and intentionally absurd. In 2026, fabrication may be fluent, disciplined, stylistically conventional, and bibliographically persuasive. The danger is therefore greater, not smaller. What once looked like nonsense may now look like scholarship.

The lesson of *Metalurgia International* is therefore not historical curiosity. It is continuity. The scholarly system has already recognized that fabricated bibliographic structures can expose a failure of editorial control. Generative AI does not invalidate that lesson.

It intensifies it.

The Indexed Paradox

The problem becomes more complex when fabricated references appear not in obscure documents circulating outside formal scholarly channels, but in articles published by journals that have passed through recognized editorial and indexing systems.

Once an article is published in an indexed journal, it is no longer merely a manuscript. It becomes part of a discoverable scholarly infrastructure. It can be searched, cited, counted, evaluated, and incorporated into institutional assessments, bibliometric analyses, promotion dossiers, grant applications, systematic reviews, and future AI training corpora.

Indexing changes the status of the text.

It does not merely reflect visibility. It amplifies it.

This is the indexed paradox. The same systems that confer legitimacy upon scholarly content may also amplify defects in that content when verification fails. If an indexed article contains nonexistent references, the issue does not remain confined to one author or one paper. It enters the broader ecology of scholarly trust.

Web of Science inclusion, in particular, carries symbolic weight. For many institutions, disciplines, and national evaluation systems, inclusion in Web of Science is treated as a proxy for quality, credibility, and international recognition. Journals included in its indexes benefit from an assumption of editorial seriousness. Their articles are more easily accepted as legitimate scholarly contributions.

But inclusion is not identity.

A journal included in Web of Science has passed a threshold at a given moment, according to a particular evaluative framework. That inclusion may be meaningful. It may indicate editorial organization, technical infrastructure, peer review practice, international relevance, and citation performance.

But it cannot guarantee the integrity of every future article.

It cannot guarantee that every reference exists.

And it cannot transform bibliographic fiction into scholarly reality.

Indexing Is Not Innocence

This editorial does not argue that indexed journals are predatory. Such a claim would be simplistic and unfair.

Nor does it argue that the presence of one fabricated reference automatically makes a journal predatory. Errors exist in all human systems. Scholarly publishing is no exception. A single incorrect reference may reflect authorial negligence, formatting error, database confusion, translation problems, careless use of reference managers, or unverified AI assistance.

Even several problematic references in one article do not automatically prove predatory conduct. They may indicate a serious failure, but not necessarily intentional deception by the journal.

The distinction matters.

Accusation is not evidence. The term “predatory” should not be used as a rhetorical weapon, nor as a substitute for analysis. It should not be applied to every weak journal, every commercial journal, every open-access journal, every fast journal, or every journal that makes an editorial mistake.

However, indexing is not innocence.

If AI-hallucinated references appear repeatedly across articles; if journals fail to detect them before publication; if editors do not correct, retract, or flag them after discovery; if publishers minimize the problem; if high-volume editorial workflows continue without adequate bibliographic verification; if indexed status is used as reputational protection against scrutiny, then the issue changes.

At that point, the question is no longer whether an error occurred.

The question is whether the editorial system is capable of distinguishing scholarly record from scholarly appearance.

This distinction becomes especially difficult when both the prose and its supporting references share the same artificial plausibility. A manuscript may read coherently, cite fluently, and follow the formal conventions of academic writing, while still lacking genuine evidentiary contact with the literature it claims to summarize. In such a case, peer review risks validating coherence rather than verification, and indexing risks amplifying appearance rather than integrity.

This is where the discussion may legitimately enter the territory of predatory-like behavior.

Not because AI hallucination itself is predatory.

Not because one contaminated article defines a journal.

But because a persistent pattern of unverifiable scholarship, combined with inadequate editorial response, begins to resemble the central feature of predatory publishing: the sale or distribution of academic validation without sufficient quality control.

Are AI-Hallucinated Citations Predatory?

The answer must be proportionate.

An AI-hallucinated reference is not automatically evidence of predatory publishing. It is, first, evidence of a verification failure.

The relevant question is the level at which that failure occurs.

At the author level, it may indicate irresponsible use of generative tools.

At the reviewer level, it may indicate that peer review did not verify the literature supporting the argument.

At the editorial level, it may indicate the absence of bibliographic due diligence.

At the publisher level, it may indicate inadequate workflow design, especially in high-volume publication systems.

At the indexing level, repeated cases may indicate that journal retention and re-evaluation policies have not yet adapted to the new risks introduced by AI-assisted writing.

Thus, the classification depends on pattern, severity, response, and persistence.

One fabricated citation may be an error.

Several fabricated citations in one article may indicate negligence.

Repeated hallucinated references across articles may indicate editorial system failure.

Failure to correct after detection may indicate post-publication irresponsibility.

Systematic tolerance of such defects, especially when combined with volume-driven publication incentives, may legitimately be described as predatory-like.

The word “like” is important. It allows us to distinguish between intentional predation and functional equivalence. A journal may not intend to deceive. But if its processes repeatedly allow unverifiable scholarship to enter the literature, and if its response is insufficient, the practical effect may still be the contamination of the scholarly record under the appearance of legitimate publication.

In such cases, bibliometric status should not protect the journal from scrutiny.

It should increase the obligation to respond.

The Problem of Metric Immunity

The difficulty is that indexed journals do not enter public suspicion in the same way as non-indexed journals.

A non-indexed journal may be suspected because of its appearance, geography, business model, website architecture, speed, language, or absence from recognized databases. A journal applying or reapplying for evaluation may be stopped at an early stage because of broken links, incomplete policies, DOI inconsistencies, metadata discrepancies, unclear contact information, or insufficiently articulated ethical statements.

Some of these observations are legitimate. Technical infrastructure matters. DOI functionality matters. Ethical policies matter. Metadata matters. Editorial transparency matters.

But if correctable technical imperfections can affect the evaluation of applicant journals, then fabricated references in indexed journals cannot be treated as minor inconveniences.

The hierarchy of seriousness must be coherent.

A broken link affects access.

A metadata discrepancy affects discoverability.

A DOI error affects traceability.

But a fabricated reference affects verifiability itself.

If evaluation systems apply forensic attention to formal infrastructure while assuming citation integrity without systematic verification, rigor becomes selective. The system may become highly sensitive to the architecture around the article, yet insufficiently attentive to the reality of the scholarly objects inside the article.

This imbalance is difficult to defend.

Bibliometric prestige may organize visibility. It may help classify journals, compare outputs, and support discovery. But it cannot function as a certificate of integrity. A journal's rank does not verify its references. An impact factor does not confirm that every cited article exists. Quartile position does not prove that peer review examined the bibliographic foundations of the manuscript.

Metrics can measure circulation.

They cannot, by themselves, certify truthfulness.

Responsibility After Publication

A common response is that citation accuracy is primarily the responsibility of authors. This is correct, but incomplete.

Authors are responsible for what they submit. They must verify that every cited source exists, that references are accurate, and that AI tools are not used to generate unsupported bibliographies.

But once a manuscript enters the publication process, responsibility becomes distributed. Reviewers evaluate the scholarly argument. Editors certify that the article has passed a process of judgment. Publishers prepare and disseminate the record. Indexing systems integrate that record into discovery and evaluation infrastructures.

Each layer contributes to the transformation of a manuscript into recognized scholarship.

A journal is not merely a container for authorial claims. It is a certifying infrastructure. It receives manuscripts, organizes peer review, applies editorial judgment, manages revisions,

prepares publication, assigns metadata, distributes content, and seeks recognition within scholarly databases.

For this reason, when fabricated references are discovered after publication, the response matters as much as the initial failure.

A journal that promptly investigates, corrects, issues an expression of concern, retracts when necessary, and revises its workflows demonstrates that editorial control still exists.

A journal that ignores, delays, minimizes, or normalizes the problem demonstrates something else.

The first failure may be detection.

The second failure may be accountability.

And the second is often more serious than the first.

The numbers also change the meaning of post-publication silence. When thousands of fabricated references can be identified across thousands of papers, and when the overwhelming majority of affected articles receive no visible publisher action at the time of audit, the issue is no longer limited to detection. It becomes a problem of response. A publication system that can discover fabricated references but does not consistently correct, flag, or retract affected records risks converting editorial failure into bibliographic permanence [2,5].

What Should Indexed Status Mean?

Inclusion in major citation indexes, including Web of Science, should mean that a journal has met defined standards of editorial quality and scholarly relevance. It should mean that the journal has passed a meaningful threshold. It should mean that its content can enter the international discovery system with a reasonable presumption of reliability.

But a presumption is not a guarantee.

Inclusion should not become immunity. It should not make scrutiny impolite. It should not transform serious editorial failures into isolated anomalies by default. And it should not create a double standard in which applicant journals are judged through formal imperfections while indexed journals are protected by reputational inertia.

If Web of Science inclusion is to retain its credibility, it must be understood as a continuing responsibility, not a permanent certificate.

This is especially important in the age of generative AI.

AI does not merely introduce new writing tools. It introduces new failure modes. The scholarly system must now distinguish between text that is fluent and text that is verified, between references that are plausible and references that exist, between bibliographic surface and bibliographic reality.

This cannot be left entirely to informal vigilance.

It requires explicit policy.

The *Metalurgia International* Question for Indexing Authorities

The *Metalurgia International* precedent leaves three deeper questions for contemporary indexing authorities.

The first concerns continuity of judgment. If fabricated bibliographic structures were once treated as evidence of serious editorial failure when they appeared in crude, visibly absurd form, should the same defect be interpreted differently when it appears in a more polished, AI-generated form? In other words, is the seriousness of fabrication determined by its visibility,

or by its nature? Does improved plausibility reduce the defect, or make it more dangerous precisely because it is harder to detect?

The second concerns the threshold between error and system failure. At what point does an AI-hallucinated reference cease to be an isolated authorial mistake and become evidence of inadequate peer review, insufficient editorial control, or defective publisher workflow? A single fabricated citation may be an error. A cluster of fabricated references may indicate negligence. Repeated cases across articles may suggest that the journal's validation process no longer reliably distinguishes scholarship from scholarly imitation.

The third concerns the responsibility of indexed status itself. If inclusion in major citation indexes, including Web of Science, confers visibility, legitimacy, and bibliometric value, should it also impose a stronger obligation of post-publication accountability when fabricated references are discovered? The question is not whether indexed journals may fail. They can. The question is whether indexing authorities will treat such failures as correctable incidents, warning signals, or grounds for re-evaluation when they affect the verifiability of the scholarly record.

The question is therefore not only qualitative, but quantitative. If fabricated references can now be detected by the thousands across millions of papers, indexing authorities can no longer treat them solely as exceptional authorial mistakes. At scale, repetition becomes evidence. The persistent appearance of fabricated references should trigger not only article-level correction, but also journal-level and publisher-level questions about workflow, verification, and accountability [2,5].

These questions matter because AI-hallucinated references cannot be managed indefinitely through informal interpretation. Without explicit criteria, responses may become selective. And selectivity is where double measure begins.

A Broader Reflection

The purpose of this editorial is not to condemn indexed journals. It is to protect the meaning of indexing.

It is not to argue that Web of Science journals are unreliable. It is to argue that reliability cannot be presumed without continuous accountability.

It is not to suggest that every AI-related error is predatory. It is to distinguish error from negligence, negligence from system failure, and system failure from predatory-like conduct.

It is not to reject artificial intelligence. It is to insist that automation without verification is incompatible with scholarly responsibility.

Generative AI will continue to develop. It will assist researchers in drafting, editing, translating, summarizing, structuring, and exploring ideas. In some contexts, it may improve clarity and accessibility. But when AI-generated references enter the scholarly record without human verification, assistance becomes contamination.

The tool has changed.

The responsibility has not.

Consistency as Integrity

Ultimately, the issue returns to consistency.

If journals outside major indexes are scrutinized for formal and technical imperfections, journals inside major indexes must also be scrutinized for failures that affect the foundations

of scholarly verification. If DOI functionality matters because it connects a text to a stable object, then reference reality matters even more because it connects a claim to the body of knowledge that supports it.

An indexed literature that cannot guarantee the reality of its references cannot fully sustain the authority of its metrics.

This does not mean that every indexed journal is suspect.

It means that no journal should be beyond scrutiny.

Indexing should open a journal to trust, not close it to examination. Ranking may organize visibility, but it cannot certify integrity. And when accusation appears, the answer should be neither automatic condemnation nor bibliometric immunity, but evidence, proportionality, correction, and the same standard applied to all.

The problem is not that indexed journals can fail.

All journals can fail.

The problem begins when failure is interpreted differently depending on status: as identity for some journals, but as an exception for others.

That is not rigor.

That is asymmetry.

And in scholarly communication, integrity cannot depend on where a journal already stands.

It must depend on what the journal continues to do.

This question remains open - and so does this series.

References

1. Zhao, Z.; Wang, Y.; Stuart, T.; De Vaan, M.; Ginsparg, P.; Yin, Y. LLM hallucinations in the wild: Large-scale evidence from non-existent citations. *arXiv preprint arXiv:2605.07723* **2026**, doi: 10.48550/arXiv.2605.07723.
2. Topaz, M.; Roguin, N.; Gupta, P.; Zhang, Z.; Peltonen, L.-M. Fabricated citations: an audit across 2.5 million biomedical papers. *The Lancet* **2026**, *407*, 1779-1781, doi:10.1016/S0140-6736(26)00603-3.
3. Naddaf, M.; Quill, E. Hallucinated citations are polluting the scientific literature. What can be done? *Nature* **2026**, *652*, 26-29, doi: 10.1080/08989621.2026.2645390.
4. Duiric, D.Z.; Delilbasic, B.; Radisic, S. Evaluation of transformative hermeneutic heuristics for processing random data. *Metal. Int* **2013**, *18*, 98-102.
5. Bauchner, H.; Rivara, F.P. Fabricated references: a new threat to editorial integrity. *Lancet* **2026**, *407*, 1765-1766, doi:10.1016/s0140-6736(26)00798-1.