

How the Imposition of AI Slop Causes Qualitative Degradation of Scholarly Output

— *An Analysis Based on the Peer-Review Process at Springer Nature* —

Date of Publication: February 23, 2026

**The details of the incident at Springer Nature discussed in this report are publicly available in our official archive and should be consulted in advance.*

1. Executive Summary

This report aims to elucidate the mechanisms by which the undisclosed use of LLMs (Large Language Models) for peer review (“AI peer review”), a practice rapidly proliferating in the modern academic publishing system, is causing qualitative degradation of scientific output. Through AI-generated-artifact forensic analysis of an incident at journals under Springer Nature (hereinafter “the subject company”), this paper goes beyond a mere whistleblowing exercise against a single corporation and introduces the concept of Coercive Model Collapse—a form of model collapse driven by asymmetric power structures. Empirical analysis has confirmed that within the subject company’s peer-review process, poor-quality AI-generated output (AI Slop) is being imposed on human authors, resulting in two concurrent phenomena: (1) a reduction in scientific diversity (Template Saturation), driven by anodyne, non-committal feedback; and (2) the circulation of critical factual errors (hallucinations), arising from the hollowing out of human oversight (Human-in-the-Loop). The subject company’s declaration of “rigorous peer review by human experts” has become effectively nominal. While collecting substantial Article Processing Charges (APCs), the system has in practice transformed into an apparatus driving a recursive contamination cycle of AI hallucinations and AI-template-mediated epistemic pollution. This report asserts that this crisis constitutes a structural defect inherent to the academic publishing system as a whole and raises an alarm that it also poses a grave threat to evaluative processes across society at large—including HR (human resources), education, and adjudication. In conclusion, it recommends a transition to “Zero-Trust Science,” independent of the legacy “peer-reviewed” brand, and the construction of a new protective framework utilizing AI forensics conducted by independent third-party organizations.

2. Introduction

This paper extends the concept of model collapse in computer science (Shumailov et al., 2024) and defines the degradation of objects mediated through power structures as “Coercive Model Collapse.” This can be broadly classified into two types. The first is the coercive pruning of diversity (Template Saturation): in the present case, this refers to the phenomenon whereby authors strip away their own originality in response to the anodyne and plausible-sounding feedback of AI reviewers, resulting in the loss of the tails of the distribution. The second is the institutional endorsement of hallucinations (the hollowing out of Human-in-the-Loop): in this case, this manifested as a situation in which the power-holding side asserted that fundamental factual errors regarding the paper’s subject matter (e.g., “GPT-5 does not exist,” “the 4o shock and hashtag campaigns are not real,” etc.) constituted “expert knowledge,” thereby authoritatively circulating falsehoods as established facts. As Naddaf (2026) demonstrates, more than half of researchers are already using AI for peer review in violation of policy, and as Chawla (2025) reports, AI detection tools are failing to capture this reality. In other words, Coercive Model Collapse is not an exceptional scandal at a single publisher but a structural crisis occurring simultaneously across the entire academic publishing system.

3. Facts of Misconduct and Cover-Up at the Subject Company

The following serious research ethics violations were confirmed in the peer-review process for a paper submitted to the subject company (submitted November 2025).

(1) Hallucinations Attributable to the LLM’s Knowledge Cutoff

In the reviewer report (Reviewer 2) for a journal specializing in AI, produced in November 2025, a series of claims were presented that would be utterly inconceivable for a human cognitive agent—such as “the events of August 2025 have not occurred” and “GPT-5 does not exist.” These are hallucinations characteristic of an LLM whose training data cutoff precedes the release of GPT-5 and whose internal search function was not activated. In this peer-review process, the company’s reviewers and editorial board evaluated only the “presentational quality of the reviewer report” and fell into a state of being unable to detect critical errors, demonstrating the complete loss of human oversight as a safety mechanism (Human-in-the-Loop).

(2) Organizational Cover-Up by Management (Laura Whitehead and the Research Integrity Division)

The subject company’s editorial board, support center, management, and Research Integrity division chose thoroughgoing dismissal and concealment in response to the author’s repeated requests for investigation. This constitutes a violation of the Committee on Publication Ethics (COPE) Core Practices. After COPE officially intervened, the editorial board released its investigation results. The head-office Research Integrity division also issued an investigation report in February 2026, but endorsed the editorial board’s findings, stating: “There is no evidence of AI generation,” “This was legitimate human judgment,” and “There was no violation of process.” What is particularly noteworthy is that, for management, this was not a case of incompetence or negligence in failing to recognize AI output (a degradation of cognitive function), but rather the manifestation of an extremely malicious managerial decision—namely, that “imposing AI Slop (poor-quality generated output) on the author was more convenient for the company than acknowledging the problem.”

Had the subject company acknowledged the reality of AI-generated peer review at this juncture, it would have triggered the following three fatal crises (a domino effect)—any of which could constitute grounds for a class-action lawsuit:

1. A clear policy violation: A self-admission that the company’s official policy of “prohibiting AI use in peer review” has been completely hollowed out.
2. A serious breach of confidentiality obligations: Exposure of the legal and ethical transgression whereby reviewers have been submitting and inputting authors’ unpublished data (submitted manuscripts) into external LLM providers without authorization.
3. Collapse of the APC (Article Processing Charge) business model: A negation of the company’s exploitative revenue structure, which charges authors thousands of dollars in high processing fees on the pretext of the added value of “rigorous and advanced judgment by human experts.”

Accordingly, the paramount imperative for management was not the pursuit of truth but “the defense of its own business model.” The methods they employed to this end were to subject the author to a barrage of dismissal, verbal abuse, and academic gaslighting (discussed below) on an organizational scale, cornering the author psychologically to force the abandonment of the complaint.

(3) Gaslighting by the Editorial Board and Head Office

Despite the author’s identification of the “AI-characteristic logical breakdown” described in (1) above, the handling editor Da Tao (Shenzhen University) and Yina Liu (Discover series) decided, following a two-month deliberate investigation, to persist in their misconduct. They dismissed the critical factual error—that “GPT-5 does not exist”—concerning the very existence of the paper’s subject matter as a “very minor element” and a “very minor aspect.” Regarding the AI hallucination of demanding “IRB approval” for data that had been categorically declared fictional while also stating that “if labeled as fictional research, it could be a valuable

contribution,” they justified it by saying “such mistakes happen occasionally at any journal.” They reported that this reviewer (i.e., the AI) “reflects careful reading and domain expertise,” that “anyone with basic training would reach the same conclusion,” and that “accepting a manuscript of this quality would raise far more serious concerns about the integrity of the review process” than rejecting it. As the editorial board itself stated, these conclusions were reached after a two-month careful investigation, and the Research Integrity division endorsed them. In other words, the editorial board and head office reported that uncritically accepting AI hallucinations was the common conclusion of any researcher with basic training, deflected the issue onto the author’s alleged lack of competence, and engaged in overt gaslighting—including dismissing the publication of empirically grounded research by a human as “inappropriate” for scientific integrity.

3. Mechanisms of Qualitative Degradation: How AI Slop Destroys Science

The case at the subject company is not an accidental occurrence but a structural defect based on the “AISP (AI Selection Pressure)” theory proposed by Naito (2025). When revision demands from AI reviewers occur in the peer-review process, qualitative degradation of scientific output is triggered through the following three stages.

3.1 Template Saturation

In computational model collapse (Early Model Collapse), the tails of the distribution are lost first. That is, minority data points are eliminated, and because the center of the distribution appears to be maintained, the collapse is not readily visible from the outside. In the context of Coercive Model Collapse, the corresponding phenomenon is “reduction of scientific diversity through Template Saturation.” AI reviewers generate a barrage of requests that are applicable to any paper but lack specificity—such as “increase the sample size” or “control for a wider range of variables.” Reviewers and editors, seeking to reduce cognitive costs, adopt these plausible-sounding suggestions as-is. As a result, peer review degenerates from a process that improves papers into the imposition of generic templates that could apply to any manuscript.

3.2 Coercive Model Collapse

Authors, whose fate is determined by acceptance decisions, are unable to mount legitimate rebuttals against AI hallucinations or off-target template-based comments.* Consequently, in order to achieve publication, researchers are compelled to pare away their own original reasoning and groundbreaking data, rewriting their papers with their own hands into “homogenized template prose that is easy for AI to learn from next.” This means that novel research occupying the tails of the distribution, minority methodologies, and interdisciplinary insights are preferentially eliminated. At present, no technical means exist to halt this collapse.

In computational Late Model Collapse, the distribution converges to a single mode, transforming into something bearing no resemblance to the original distribution. In the context of Coercive Model Collapse, the corresponding phenomenon is the complete loss of human oversight safety (Human-in-the-Loop). In this collapse, AI hallucinations circulate directly as scientific facts. In the subject company’s case, factual errors of a caliber that would never occur in a peer review for an AI-specialized journal—such as “GPT-5 does not exist”—were generated. More broadly, erroneous suggestions at a level that negates modern science itself were produced in practice—such as a “human expert” (actually an AI reviewer) denying the existence of a paper’s actual research subject and then proposing that “it could still have alternative value if labeled as fiction.” All of these originate from AI hallucinations and are qualitatively distinct from errors made by human reviewers.

3.3 Completion of the Recursive Pollution Cycle

After undergoing the above process several times, papers revised by authors to conform to AI demands are published under brands such as Nature as “peer-reviewed papers.” These are then ingested as training data for the next generation of AI, causing the diversity (variance) of science to contract, and templates and falsehoods become entrenched as scientific facts. While the “gradual reduction of diversity” is likely to become a long-term problem, the undeniably urgent issue at present is that “short-term, critical AI errors” are being allowed to pass unchallenged in practice. The former gradually erodes the foundations of science, but the latter

produces immediately critical consequences. It is not difficult to imagine that in fields such as medicine, engineering, and pharmacology, if a paper containing AI hallucinations that passed through peer review is applied to practice under the authority of the “peer-reviewed” label, the impact will directly affect human lives. This recursive contamination cycle driven by AI peer review functions as a feedback loop that amplifies both types of problems through authors’ (human) learning and preemptive adaptation to AI-based checks.

****Unable to mount legitimate rebuttals***

The subject company ostensibly maintains “Speak Up” (internal whistleblowing/consultation) channels and an “Appeal” process, presenting itself as transparent and cooperative toward authors’ legitimate objections. In reality, however, even when authors use these official channels to report clear-cut misconduct such as AI-generated peer review, the following “defense protocol” completely neutralizes such efforts:

a. Requirement for External Pressure (COPE) and Outright Dismissal

Even when an author independently presents clear evidence (such as logs of AI hallucinations), the Research Integrity division refuses to even issue a case number, maintaining a stance of investigation refusal. An investigation is only initiated when formal intervention from an external body such as COPE (Committee on Publication Ethics) occurs.

b. Dysfunction of the Speak Up (Whistleblowing) System

Even when a serious ethics violation is reported to the compliance channel, its contents are forwarded not to an independent third party but directly back to the editorial board under complaint (i.e., the perpetrator of the misconduct). As a result, even issues of systemic misconduct are forcibly reframed as a mere “routine appeal” of a decision by the author, and the author’s will to fight is worn down through demands for meaningless additional data and delays of several weeks.

c. Hollowing Out of Management and Circular Logic

Most egregiously, the logic of the investigation itself has collapsed. Publisher-side and Research Integrity division management do not conduct objective forensics or independent investigations; instead, they simply endorse the self-reporting of the parties involved, declaring: “The editorial board responded that there was no violation; therefore the process is legitimate.” Furthermore, even when presented with evidence of AI-characteristic factual errors, the editorial side deploys anti-scientific circular reasoning such as: “Our peer review has an established policy on AI use; therefore there is no evidence that this review was wholly delegated to AI,” and “The peer-review process and its outcome were legitimate, but we will continue to investigate whether there was an AI policy violation”—statements so incoherent as to suggest a failure to comprehend the very meaning of an investigation.

In conclusion, no matter how much scientific and objective evidence (legitimate rebuttal) the author presents, it is crushed through denial of facts. This is the primary reason why, under power structures, authors are forced to accept AI Slop and have no choice but to acquiesce (i.e., to accept Coercive Model Collapse). This protocol at the review level has been enforced down to the assistant editor level, where it was emphasized that objections would be deprioritized below new submissions and that decisions would take several weeks, and that “in the majority of cases the original decision will be upheld.” In other words, by implicitly sending the message that objecting is futile and a mere waste of time, they induce the author to accept the situation as-is—that “once the editorial board has officially reported a result, it will not be overturned even if there are clear problems.”

4. Conclusion and Practical Remediation

Conclusion

4-0 Peer review is not a warranty mechanism but a weak quality signal

The subject company’s peer-review system has metamorphosed into an apparatus that overwrites truth with AI-generated falsehoods (hallucinations). This case has demonstrated that the de facto tolerance of AI peer review leads not merely to a decline in quality but directly to academic harassment and the coercion of data fabrication. The structure of Coercive Model Collapse, as articulated in the existing literature and in this paper, indicates that this crisis is not the scandal of a single publisher but a structural

crisis inherent to the academic publishing system as a whole. Accordingly, the conscientious scientific community and relevant institutions should execute the following transition to “Zero-Trust Science.”*

*Zero-Trust Science

While it is neither feasible nor desirable to designate every preprint as a “trusted preprint,” the company recognizes the necessity of targeted assurance in high-impact contexts. In particular, when research outputs are used in funding allocation, medical decision-making, regulatory evaluation, or disputes arising within the peer-review process, stakeholders should actively employ Disclosure of peer-review system logs (including AI usage traces and prompt histories) and Independent third-party audits(including AI-generated content forensics,).

Rather than attaching uniform guarantees at the point of publication, assurance mechanisms should be event-triggered, escalating in proportion to real-world impact and risk. This approach preserves the openness of preprint culture while preventing low-quality AI-generated content from acquiring unwarranted authority in consequential decision environments.

4-1 Invalidation of the “Peer-Reviewed” Brand

A shift should be made toward a standard that places greater value on raw data (preprints) from trustworthy authors and organizations as primary sources, rather than peer-reviewed papers that have been subjected to “censorship and falsification” by AI and dishonest editors behind closed doors. The authority of the “peer-reviewed” label conferred by academic publishers whose AI peer-review policies are in routine violation, or whose policies and practices regarding AI peer review remain opaque, should not be trusted. Peer review is not a warranty mechanism but a weak quality signal.

4-2 Active Utilization of “AI-Generated Content Forensics” Services by Third-Party Organizations

Using AI-generated content forensics by human experts provided by UTIE Instruments Inc. and other, protocols for technically and objectively auditing whether the content of a paper contains “traces of low-quality AI-generated manuscripts (chronological inconsistencies, template-like convergence, hallucinations)” should be incorporated into criteria for the allocation of public research funding and similar frameworks.

4-3 Relief from “Coercive Assimilation by Power” and Construction of Ethical Protective Frameworks

The crisis demonstrated by this case is less a defect of AI itself than a structure in which “the party holding power delegates work wholesale to AI, and then coercively assimilates the results (hallucinations and template conformity) onto those in weaker positions.” In order to protect humans from this “Coercive Model Collapse,” the following frameworks must be urgently constructed:

4-3-1 Intervention by Third-Party Organizations (e.g., COPE) and Guideline Development in Academia

It has become clear that the internal “Research Integrity divisions” of publishers—which prioritize their own interests (high APC models and cost reduction through AI)—possess neither the capacity for self-correction nor the ability to conduct objective investigations. Accordingly, the system should transition to one in which independent third-party organizations such as COPE (Committee on Publication Ethics) take the lead in managing and intervening in matters such as accepting objections to undue wholesale-AI-delegated peer review and requiring disclosure of system logs from the peer-review process (traces of AI use, prompt histories, etc.). In order to spare authors the enormous effort of disputes with corporations, the development of “Transparency and Relief Guidelines for AI-Delegated Peer Review” is urgently needed.

4-3-2 Protection from “Unfair AI Evaluation” Across Society (HR, Education, and Adjudication)

The imposition of low-quality AI-generated output through asymmetric power structures is not a problem confined to academic publishing. Currently, on a global and simultaneous basis, the delegation of judgments to AI by evaluators seeking cost reduction is increasing across all evaluative processes that shape human lives—including job interviews, internal personnel evaluations, and the adjudication of grants and financing. As a result, individuals possessing unique strengths or groundbreaking potential can be deemed “unqualified” and eliminated by AI screening simply because they deviate from the AI’s training data (i.e., historical

averages). A safety net must be urgently constructed across society as a whole, enabling those subject to evaluation to demand disclosure of the grounds for judgment and re-examination by human experts in the face of black-boxed AI determinations (and the humans or organizations that lack the ability to verify their validity).

References

- Chawla, D. S. (2025). “‘A serious problem’: peer reviews created using AI can avoid detection.” *Nature*.
- Naddaf, M. (2026). “More than half of researchers now use AI for peer review — often against guidance.” *Nature*.
- Naito, H. (2025). “AI Selection Pressure: Template Saturation and the Reshaping of Human Discernment.” Zenodo. <https://doi.org/10.5281/zenodo.17644956>
- Shumailov, I., Shumaylov, Z., Zhao, Y., Gal, Y., Papernot, N., & Anderson, R. (2024). “AI models collapse when trained on recursively generated data.” *Nature*, 631, 755–759.

© 2026 UTIE Instruments Inc. All rights reserved. No part of this publication may be reproduced without identifying the source as UTIE Instruments Inc.