SWISS DENTAL JOURNAL SSO

Science and Clinical Topics

ISSN 2296-6498
swissdentaljournal.org

Scientific article

Who reviewed this? Toward responsible integration of large language models for peer review of scientific articles in dental medicine

Accepted: September 5, 2025
DOI: 10.61872/sdj-2025-03-01
2025, Vol. 135

Florin Eggmann[1*], Hauke Hildebrand[1], Michael M. Bornstein[2]

[1] Department of Periodontology, Endodontology, and Cariology, University Center for Dental Medicine Basel UZB, University of Basel, Basel, Switzerland
[2] Department of Oral Health & Medicine, University Center for Dental Medicine Basel UZB, University of Basel, Basel, Switzerland

*Correspondence: Florin Eggmann, Department of Periodontology, Endodontology, and Cariology, University Center for Dental Medicine Basel UZB, University of Basel, Mattenstr. 40, CH-4058 Basel, Switzerland.
Telephone number: +41 61 267 26 80, email: florin.eggmann@unibas.ch

Keywords

Artificial intelligence, Dentistry, Editorial Policies, Machine Learning, Peer Review Research, Scholarly Communication

Abstract

The introduction and advancement of large language models (LLMs), such as ChatGPT, DeepSeek, and Google Gemini, present both opportunities and challenges for peer review in dental research. In this article, we propose a framework to inform the discourse on the responsible use of LLMs in dental peer review. We conducted a cross-sectional review of peer review policies from the top 50 dental journals, based on their 2024 Journal Impact Factor, to assess current guidance on LLM use. Our analysis revealed variability across dental journals: some journals permit restricted LLM use under specific conditions, while many either prohibit their use or lack explicit policies. Key concerns regarding LLM use identified by the authors include potential breaches of confidentiality, ambiguity in authorship, reduced reviewer accountability, and inherent limitations of LLMs in terms of domain-specific expertise and factual accuracy. Our proposed framework addresses confidentiality safeguards, suggested appropriate LLM applications, areas requiring caution, disclosure requirements, and accountability standards. It emphasizes that reviewers retain full responsibility for all submitted content, irrespective of LLM assistance. To protect confidentiality, the framework encourages offline or locally hosted LLMs. It also recommends regular policy reviews and reviewer training. This framework aims to support the thoughtful adoption of LLMs in dental research publishing. When employed judiciously, LLMs offer potential benefits in improving review clarity and efficiency, particularly for reviewers writing in a non-native language. However, their use must be grounded in clear ethical principles to ensure the integrity of dental peer review.

Introduction

Large language models (LLMs) are generative artificial intelligence systems trained on vast corpora of text. These models analyze the statistical relationships between tokens, the basic units of language used in LLM training, which can include words, word fragments, punctuation marks, or even individual characters (1). By predicting the most likely sequence of tokens in response to any given prompt, LLMs can generate coherent, contextually appropriate, and human-like text (1-3).

Commercially available chatbots underpinned by LLMs, including ChatGPT, Claude, DeepSeek, and Google Gemini, have demonstrated notable capabilities across a wide range of domains, including dental medicine (2-4). They are commonly used for tasks such as summarization, translation, answering questions, and drafting emails, reports, and even computer code (1). Despite lacking true understanding or reasoning, LLMs can simulate human-like dialogue and writing with increasing sophistication (1-3).

As LLMs become more integrated into professional and academic workflows, their use raises questions about authorship and scientific integrity (5, 6). One area of interest is the peer review process (7-18). Evidence from textual analyses and anecdotal reports suggest that some reviewers use these tools to help structure, summarize, or even draft their reviews (19, 20).

LLMs may offer certain advantages, particularly for reviewers whose first language is not English, the prevailing lingua franca of academic publishing (1, 21). Moreover, assistance from LLMs has demonstrated utility in enhancing clarity of communication, including in patient interactions (22, 23). The benefit of clearer communication may also extend to peer review because improved clarity can shape how information is received and acted upon.

However, the unregulated and undisclosed use of LLMs raises concerns regarding the transparency and quality of peer review (1, 11). For instance, uploading unpublished manuscripts to online platforms may breach confidentiality and compromise intellectual property (14). In addition, LLM outputs are prone to factual inaccuracies and lack clear accountability (1). These issues call into question reviewer responsibility, authorship, and transparency, and risk undermining the expert assessment, cogent feedback, and critical judgment that peer reviewers are intended to provide.

In this article, we examine peer review policies of top-ranked dental journals and propose a framework of principles and practices to guide the responsible use of LLMs in peer review in dental medicine. Our goal is to contribute to the ongoing discourse on trust and ethical standards in dental research publishing in the age of artificial intelligence.

Materials and methods

We conducted a cross-sectional review of the peer review policies of the top 50 dental journals, as ranked by the 2024 Journal Impact Factor (JIF), to assess current guidance on the use of LLMs in the peer review process. The list of journals was obtained from the 2024 Journal Citation Reports (Clarivate Analytics, Philadelphia, PA, USA), using the subject category “Dentistry, Oral Surgery & Medicine.”

Between July 22 and July 24, 2025, we accessed the official websites of each journal and systematically reviewed all publicly available peer review policy information. We examined peer review and editorial policies. When explicit policies were not available on a journal’s site, we reviewed the associated publisher’s general editorial or ethics policies (e.g., Elsevier, Springer Nature, Wiley).

Data extraction included review model, confidentiality of peer review, explicit mention of artificial intelligence, allowance or prohibition of LLMs, LLM use disclosure requirements, and accountability clauses. Two authors independently reviewed each policy. Any discrepancies were resolved through discussion (between F. E. and H. H.).

Through team discussion among all three authors, we compiled a list of potential risks associated with LLM use in peer review. The findings of our review, together with these identified risks, informed the development of a proposed framework for responsible LLM use in peer review. This framework was drafted collaboratively, iteratively refined, and finalized by consensus among all authors.

Table 1. Peer review policy overview of the 50 top-ranked dental journals

Journal (publisher)	JIF 2024	Review model	Confiden-tiality	AI mentioned	LLM allowed (permitted with condi-tions) / Pro-hibited / Un-clear	LLM usage disclosure required	Accounta-bility	Guideline link (access date)
Periodontol-ogy 2000 (Wiley)	15.7	Single-blind	Yes	Yes	Permitted with conditions (en-hance written feedback)	Yes	Yes	https://authorserv-ices.wiley.com/eth-ics-guidelines/in-dex.html#15 (July 24, 2025)
International Journal of Oral Science (Springer Na-ture)	12.2	Double-blind	Yes	Yes	Permitted with conditions (ex-plicitly prohib-iting the up-loading of manuscripts to LLMs)	Yes	Yes	https://www.springeropen.com/get-pub-lished/editorial-poli-cies#artificial+intelli-gence+%28ai%29 (July 24, 2025)
International Endodontic Journal (Wiley)	7.1	Double-blind	Yes	Yes	Permitted with conditions (en-hance written feedback)	Yes	Yes	https://authorserv-ices.wiley.com/eth-ics-guidelines/in-dex.html#15 (July 24, 2025)
Journal of Clinical Perio-dontology (Wiley)	6.8	Double-blind	Yes	Yes	Permitted with conditions (en-hance written feedback)	Yes	Yes	https://authorserv-ices.wiley.com/eth-ics-guidelines/in-dex.html#15 (July 24, 2025)
Japanese Den-tal Science Re-view (Elsevier)	6.6	Double-blind	Yes	Yes	Prohibited	N/A	Yes	https://www.else-vier.com/re-viewer/how-to-re-view (July 24, 2025)
Dental Mate-rials (Elsevier)	6.3	Double-blind	Yes	Yes	Prohibited	N/A	Yes	https://www.else-vier.com/re-viewer/how-to-re-view (July 24, 2025)
Journal of Dental Re-search (Sage)	5.9	Single-blind	Yes	Yes	Permitted with conditions (en-hance written feedback)	No	Yes	https://www.sagepub.com/journals/edi-torial-policies/artifi-cial-intelligence-pol-icy (July 24, 2025)
Journal of Dentistry (Elsevier)	5.5	Double-blind	Yes	Yes	Prohibited	N/A	Yes	https://www.else-vier.com/re-viewer/how-to-re-view (July 24, 2025)
Clinical Oral Implants Re-search (Wiley)	5.3	Single-blind	Yes	Yes	Permitted with conditions (en-hance written feedback)	Yes	Yes	https://authorserv-ices.wiley.com/eth-ics-guidelines/in-dex.html#15 (July 24, 2025)

Progress in Orthodontics (Springer Nature)	5.0	Double-blind	Yes	Yes	Permitted with conditions (explicitly prohibiting the uploading of manuscripts to LLMs)	Yes	Yes	https://www.springeropen.com/get-published/editorial-policies#artificial+intelligence+%28ai%29 (July 24, 2025)
Journal of Prosthetic Dentistry (Elsevier)	4.8	Double-blind	Yes	Yes	Prohibited	N/A	Yes	https://www.elsevier.com/reviewer/how-to-review (July 24, 2025)
Journal of Esthetic and Restorative Dentistry (Wiley)	4.1	Double-blind	Yes	Yes	Permitted with conditions (enhance written feedback)	Yes	Yes	https://authorservices.wiley.com/ethics-guidelines/index.html#15 (July 24, 2025)
Dentomaxillofacial Radiology (Oxford Academic)	4.1	Double-blind	Yes	Yes	Permitted with conditions (prohibition on uploading manuscript content, using LLMs as expert opinion substitutes, or superseding reviewer judgment)	No	Yes	https://academic.oup.com/dmfr/pages/reviewer-guidelines (July 24, 2025)
Clinical Implant Dentistry and Related Research (Wiley)	4.0	Single-blind	Yes	Yes	Permitted with conditions (enhance written feedback)	Yes	Yes	https://authorservices.wiley.com/ethics-guidelines/index.html#15 (July 24, 2025)
Journal of Oral Rehabilitation (Wiley)	4.0	Single-blind	Yes	Yes	Permitted with conditions (enhance written feedback)	Yes	Yes	https://authorservices.wiley.com/ethics-guidelines/index.html#15 (July 24, 2025)
Journal of Evidence-Based Dental Practice (Elsevier)	4.0	Single-blind	Yes	Yes	Prohibited	N/A	Yes	https://www.elsevier.com/reviewer/how-to-review (July 24, 2025)
International Journal of Implant Dentistry (Springer Nature)	4.0	Single-blind	Yes	Yes	Permitted with conditions (explicitly prohibiting the uploading of manuscripts to LLMs)	Yes	Yes	https://www.springeropen.com/get-published/editorial-policies#artificial+intelligence+%28ai%29 (July 24, 2025)
Dental and Medical Problems (Wroclaw Medical University Press)	3.9	Double-blind	Yes	Yes	Prohibited	N/A	Yes	https://dmp.umw.edu.pl/en/review-policy/ (July 24, 2025)

Oral Oncology (Elsevier)	3.9	Double-blind	Yes	Yes	Prohibited	N/A	Yes	https://www.elsevier.com/reviewer/how-to-review (July 24, 2025)
Journal of Periodontology (Wiley)	3.8	Single-blind	Yes	Yes	Permitted with conditions (enhance written feedback)	Yes	Yes	https://authorservices.wiley.com/Reviewers/journal-reviewers/tools-and-resources/review-confidentiality-policy.html (July 24, 2025)
International Dental Journal (Elsevier)	3.7	Single-blind	Yes	Yes	Prohibited	N/A	Yes	https://www.elsevier.com/reviewer/how-to-review (July 24, 2025)
Journal of Endodontics (Elsevier)	3.6	N/A	Yes	Yes	Prohibited	N/A	Yes	https://www.elsevier.com/reviewer/how-to-review (July 24, 2025)
Journal of Prosthodontics (Implant Esthetic and Reconstructive Dentistry) (Wiley)	3.6	Double-blind	Yes	Yes	Permitted with conditions (enhance written feedback)	Yes	Yes	https://authorservices.wiley.com/Reviewers/journal-reviewers/tools-and-resources/review-confidentiality-policy.html (July 24, 2025)
Journal of the American Dental Association (Elsevier)	3.5	Double-blind	Yes	Yes	Prohibited	N/A	Yes	https://www.elsevier.com/reviewer/how-to-review (July 24, 2025)
Journal of Prosthodontic Research (Japan Prosthodontic Society)	3.4	Double-blind	Not specified	No	Not specified	N/A	Not specified	https://hotetsu.com/eng/index.html (July 24, 2025)
Journal of Periodontal Research (Wiley)	3.4	Single-blind	Yes	Yes	Permitted with conditions (enhance written feedback)	Yes	Yes	https://authorservices.wiley.com/ethics-guidelines/index.html#15 (July 24, 2025)
Angle Orthodontist (The E. H. Angle Education and Research Foundation)	3.2	Double-blind	Not specified	No	Not specified	N/A	Not specified	https://angle-orthodontist.kglmeridian.com/page/how_to_submit (July 24, 2025)
Journal of the World Federation of Orthodontists (Elsevier)	3.2	Double-blind	Yes	Yes	Prohibited	N/A	Yes	https://www.elsevier.com/reviewer/how-to-review (July 24, 2025)

Journal of Periodontal and Implant Science (Korean Academy of Periodontology)	3.1	Not specified	Not specified	No	Not specified	N/A	Not specified	https://jpis.org (July 24, 2025)
Journal of Dental Sciences (Elsevier)	3.1	Double-blind	Yes	Yes	Prohibited	N/A	Yes	https://www.elsevier.com/reviewer/how-to-review (July 24, 2025)
BMC Oral Health (Springer Nature)	3.1	Single-blind	Yes	Yes	Permitted with conditions (explicitly prohibiting the uploading of manuscripts to LLMs)	Yes	Yes	https://www.springeropen.com/get-published/editorial-policies#artificial+intelligence+%28ai%29 (July 24, 2025)
Clinical Oral Investigations (Springer Nature)	3.1	Single-blind	Yes	Yes	Permitted with conditions (explicitly prohibiting the uploading of manuscripts to LLMs)	Yes	Yes	https://www.springeropen.com/get-published/editorial-policies#artificial+intelligence+%28ai%29 (July 24, 2025)
Dental Traumatology (Wiley)	3.1	Double-blind	Yes	Yes	Permitted with conditions (enhance written feedback)	Yes	Yes	https://authorservices.wiley.com/ethics-guidelines/index.html#15 (July 24, 2025)
Dentistry Journal (MDPI)	3.1	Single-blind (Open review available)	Yes	Yes	Permitted with conditions (enhance written feedback)	Yes	Yes	https://www.mdpi.com/reviewers (July 24, 2025)
Frontiers in Oral Health (Frontiers Media SA)	3.1	Single-blind	Not specified	Yes	Permitted with conditions	Yes	Yes	https://www.frontiersin.org/guidelines/policies-and-publication-ethics?utm_source=chatgpt.com (July 24, 2025)
American Journal of Orthodontics and Dentofacial Orthopedics (Elsevier)	3.0	Double-blind	Yes	Yes	Prohibited	N/A	Yes	https://www.elsevier.com/reviewer/how-to-review (July 24, 2025)
Oral Diseases (Wiley)	2.9	Single-blind	Yes	Yes	Permitted with conditions (enhance written feedback)	Yes	Yes	https://authorservices.wiley.com/ethics-guidelines/index.html#15 (July 24, 2025)
Molecular Oral Microbiology (Wiley)	2.9	Single-blind	Yes	Yes	Permitted with conditions (enhance written feedback)	Yes	Yes	https://authorservices.wiley.com/ethics-guidelines/index.html#15 (July 24, 2025)

Maxillofacial Plastic and Reconstructive Surgery (Springer Nature)	2.8	Double-blind	Yes	Yes	Permitted with conditions (explicitly prohibiting the uploading of manuscripts to LLMs)	Yes	Not specified	https://www.springeropen.com/get-published/editorial-policies#artificial+intelligence+%28ai%29 (July 24, 2025)
European Journal of Paediatric Dentistry (Ariesdue Srl)	2.7	Not specified	Not specified	No	Not specified	N/A	Yes	https://www.ejpd.eu/guidelines-for-authors/ (July 24, 2025)
International Journal of Oral and Maxillofacial Surgery (Elsevier)	2.7	Not specified	Yes	Yes	Prohibited	N/A	Yes	https://www.elsevier.com/reviewer/how-to-review (July 24, 2025)
European Journal of Orthodontics (Oxford University Press)	2.7	Double-blind	Yes	Yes	Permitted with conditions	Yes	Yes	https://publicationethics.org/guidance/guideline/ethical-guidelines-peer-reviewers (July 24, 2025)
Journal of Adhesive Dentistry (Quintessence Publishing)	2.7	Double-blind	Yes	Yes	Prohibited	N/A	Yes	https://www.quintessence-publishing.com/downloads/ai-use-policy_qi.pdf (July 24, 2025)
Head & Face Medicine (Springer Nature)	2.6	Open peer review	Yes	Yes	Permitted with conditions (explicitly prohibiting the uploading of manuscripts to LLMs)	Yes	Yes	https://www.springeropen.com/get-published/editorial-policies#artificial+intelligence+%28ai%29 (July 24, 2025)
Caries Research (S. Karger AG)	2.6	Single-blind	Yes	Yes	Prohibited	N/A	Yes	https://karger.com/DocumentLibrary/Reviewer_Guidelines_Karger.pdf (July 24, 2025)
Journal of Applied Oral Science (University of São Paulo)	2.6	Single-blind (Open review available)	No	No	Not specified	N/A	Yes	https://www.scielo.br/journal/jaos/about/#instructions (July 24, 2025)
Journal of Oral and Maxillofacial Surgery (Elsevier)	2.6	Not specified	Yes	Yes	Prohibited	N/A	Yes	https://www.elsevier.com/reviewer/how-to-review (July 24, 2025)
Journal of Advanced Prosthodontics (Korean Academy of Prosthodontics)	2.5	Single-blind	No	No	Prohibited	N/A	Not specified	https://jap.or.kr/src/JAP-Instructions-to-Authors-202408.pdf (July 24, 2025)

International Journal of Oral Implantology (Quintessence Publishing)	2.5	Double-blind	Yes	Yes	Prohibited	N/A	Yes	https://www.quintessence-publishing.com/downloads/ai-use-policy_qi.pdf (July 24, 2025)
Odontology (Springer Nature)	2.4	Single-blind	Yes	Yes	Permitted with conditions (explicitly prohibiting the uploading of manuscripts to LLMs)	Yes	Yes	https://www.springeropen.com/get-published/editorial-policies#artificial+intelligence+%28ai%29 (July 24, 2025)

Abbreviations: AI (artificial intelligence), JIF (Journal Impact Factor), LLM (large language model)

Results

Overview of dental journal polices

Table 1 presents an overview of journal policies of the 50 top-ranked dental journals.

Among the 50 journals, 43 (86%) explicitly mentioned artificial intelligence in their reviewer or ethics policies. Their policy positions regarding peer reviewers' use of LLMs were distributed as follows:

- 26 journals (52%) permitted LLM use with specified conditions.
- 19 journals (38%) prohibited LLM use outright.
- 5 journals (10%) had either unclear or no explicit policies.

LLM use, where permitted, was universally conditional—typically limited to providing linguistic feedback, enhancing clarity, or explicitly prohibiting the uploading of manuscripts to LLMs. Crucially, many journals emphasized that LLMs must not be used for evaluative judgment, content assessment, or as a substitute for reviewer expertise.

Disclosure and accountability requirements

Among the 26 dental journals permitting conditional use of LLMs, 24 (92%) required explicit disclosure of LLM use during the review process. In contrast, none of the journals that prohibited LLMs provided disclosure pathways, which implies a zero-tolerance policy. Accountability for the review remained with the human reviewer across journals, irrespective of LLM involvement.

Potential risks of LLM use in peer review

Through team discussion, we identified the following risks of LLM use in peer review:

- **Violations of journal policy**: Most dental journals explicitly prohibit or restrict LLM use in peer review. Using LLM tools without adherence to stated guidelines may constitute a breach of editorial policy.
- **Confidentiality breaches**: Most peer review systems operate under strict confidentiality. Uploading manuscripts, wholly or in part, into LLM platforms may violate author privacy, publisher policies, and potentially intellectual property rights. Moreover, many commercial LLMs retain user-inputted data to train future models. Even when anonymized, such data could contribute to indirect leakage of proprietary or sensitive information.
- **Prompt injection by authors**: Anecdotal reports suggest that authors may embed concealed instructions, such as white-text prompts, within manuscripts to manipulate LLM-assisted reviewers into generating favorable evaluations (24). This represents a concerning avenue for misuse.
- **Opaque authorship of reviews**: Using LLMs may obscure the human contribution to peer reviews. This lack of transparency may undermine reviewer accountability, complicate questions of authorship, and cast doubt on the reliability and expertise of the reviewer.
- **Inaccuracy and hallucination**: LLMs are prone to fabricating references, misinterpreting scientific content, and generating plausible-sounding yet incorrect critiques. Such errors have the potential to mislead reviewers, editors, and authors alike (1, 23).

- **Bias and lack of domain expertise**: LLMs are trained on general corpora of textual data and lack the nuanced reasoning and critical judgment essential to high-quality peer review (25). Their use risks introducing systemic biases or diluting the domain-specific rigor (13).
- **Undisclosed use and editorial blind spots**: Currently, there are no standardized mechanisms requiring reviewers to disclose LLM use (13). This lack of transparency deprives editors and journals of crucial information about how review content is generated.
- **Decline in mentorship and skill development**: Peer review often serves as a skill-building opportunity for early-career researchers through co-review with senior colleagues. Increased reliance on LLMs may reduce opportunities for mentorship and collaborative learning. This, in turn, could lead to a decline in critical appraisal skills within the dental community, particularly among early-career researchers.

Proposed framework for LLM use in peer review

To ensure high standards in the peer review process in the age of LLM chatbots, we propose the following framework of principles and practices, which we recommend dental journals adapt and implement accordingly:

Reviewer accountability

Reviewers remain fully accountable for the accuracy, fairness, and scientific integrity of their submitted review, regardless of whether an LLM was used during its preparation. Reviewers should not use LLMs to compensate for a lack of subject-matter expertise. Any text generated with the aid of an LLM should be carefully reviewed, appropriately revised, and verified by the reviewer to ensure it meets the expected evaluation standards.

Confidentiality and data security

Confidentiality and data security are central to the integrity of the peer review process. Reviewers should exercise great caution when using LLMs, particularly with respect to unpublished material. Uploading any part of a non-public manuscript, including text, data, tables, or figures, to online LLM platforms not covered by institutional data-use agreements entails serious risks to confidentiality, especially if such services do not guarantee data non-retention.

Where LLM use is considered appropriate, locally hosted or offline models, such as GPT4ALL (Nomic AI, New York, NY, USA), may offer a safer alternative by avoiding transmission of sensitive content to external servers. Step-by-step guidance is available to help reviewers use such tools (26).

Disclosure and transparency

Reviewers should disclose any use of LLMs in the preparation of their review, including the specific purpose, such as summarization, grammar correction, or structural editing. To support transparency, journals could consider incorporating a standard checkbox and an accompanying comment field into their review submission platforms to document this information.

Restrictions on LLM use

The use of LLMs in peer review should be limited to clearly defined, low-risk tasks. Permissible uses may include language refinement, grammar correction, and improving the clarity or structure of a reviewer’s own scientific critique. These applications may be particularly helpful for enhancing readability and expression.

In contrast, reviewers should not delegate core evaluative functions, such as assessing scientific merit, methodological rigor, or the novelty of a manuscript, to an LLM. Similarly, it is inappropriate to use LLMs to summarize manuscripts that are not in the public domain or to generate or suggest review scores and editorial recommendations, such as decisions to accept, revise, or reject.

Monitoring and policy enforcement

As LLMs and related tools continue to advance in capability and reach, their use in dental peer review warrants regular critical evaluation. Editors, publishers, and other stakeholders, both within and beyond the dental community, play an important role in maintaining over-sight to support the responsible, transparent, and ethical integration of machine learning technologies. Instances where LLMs are used in ways that contravene journal guidelines may call for appropriate remedial action.

Training and education

Dental journals and research institutions are encouraged to provide guidance and training on the ethical and responsible use of LLMs in peer review. As these tools become more widely adopted, proactive education can help reviewers understand appropriate boundaries and limitations. In parallel, editorial bodies may benefit from working collaboratively to develop shared policies and reviewer training materials. Establishing consistent expectations across dental journals would support greater transparency in the use of LLMs throughout the peer review process.

Moreover, it is crucial to protect opportunities for mentorship and skill development. Early-career researchers should continue to engage in the formative aspects of peer review, which are best supported through direct collaboration with senior colleagues.

Discussion

The integration of LLMs into the peer review process holds promise for enhancing clarity, efficiency, and inclusivity, particularly for reviewers working in a non-native language. However, their use also introduces ethical and procedural concerns. Key concerns include transparency, data protection, and human accountability.

Our review of the 50 top-ranked dental journals revealed marked variability in existing editorial policies concerning LLM use. While several journals explicitly prohibit their use, others allow it under specified conditions—most commonly for linguistic refinement—and some provide no guidance at all. The way journals address LLM use also differs: some offer clear instructions for reviewers, whereas others only reference policies directed at authors that include brief mentions relevant to reviewers. Moreover, many journals refer to broader standards issued by organizations such as the International Association of Scientific, Technical, and Medical Publishers (STM) or the Committee on Publication Ethics (COPE). Yet, these external frameworks themselves vary in clarity and scope regarding LLMs (11). This inconsistency highlights two challenges. First, it places the onus on individual reviewers to interpret and comply with divergent or ambiguous policies. This, in turn, entails risks of inadvertent non-compliance. Second, it underscores the need for harmonized, context-sensitive, and regularly updated guidelines that reflect the changing capabilities and limitations of machine learning technologies.

Our findings in dental journals align with a recent survey of medical journals, which also revealed wide variation in policies on LLM use in peer review (16). In both fields, most journals either prohibit such use outright or allow it only under strict conditions, with confidentiality emerging as the primary concern (16). Where permitted, LLMs are generally limited to language editing and must be disclosed (16). This consistency across disciplines underscores the common risks involved and the need for clearer, harmonized guidelines.

Developing actionable guidelines for LLM-assisted scientific writing is difficult. One helpful step is the taxonomy introduced by STM, which defines key terms related to the use of machine learning tools in academic publishing (27). These definitions provide a common language that should be reflected in any formal guidelines for LLM-assisted writing and peer review. Without a shared understanding of key terms, it is difficult to develop policies that are both clear and practical.

Recent work on frameworks such as the Generative AI Delegation Taxonomy (GAIDeT) offers a useful way to think about and document LLM use in peer review (17). GAIDeT breaks down the contributions of machine learning tools, including LLMs, across different research tasks and emphasizes clear human oversight and accountability (17). These principles align with the cautious approaches seen in most dental journals. Using such a framework could help reviewers and editors clearly record when and how LLMs are used.

The framework proposed in this paper is intended to be neither prescriptive nor exhaustive. Rather, it aims to contribute to the ongoing discourse on how LLMs might be integrated into the peer review process in ways that uphold scientific integrity (14-18). We advocate for a cautious, evidence-informed approach with regular reassessment. Such an approach may enable the dental research community to responsibly realize the benefits of LLMs—enhancing the clarity, efficiency, and accessibility of peer reviews—especially for reviewers for whom English is a second language.

Offline or locally hosted models offer promising solutions to mitigate confidentiality risks. Judicious use of such LLM tools can enhance reviewer efficiency and support reviewers in producing clear, well-structured reports. For example, a reviewer might focus on assessing the scientific content of a manuscript, then use an offline LLM to transform notes or dictated observations into coherent, section-by-section feedback (21). Offline models can also assist in polishing written feedback, while ensuring that reviewers retain control over content and confidentiality (21). Besides locally hosted models, some publishers, such as Elsevier, are developing their own LLM tools to assist authors and reviewers. These LLM tools are designed to work within established policy frameworks, helping users comply with guidelines while benefiting from LLM assistance (5).

These developments unfold against the backdrop of a peer review system that is already under strain. The increasing burden on reviewers, coupled with a frequent lack of formal recognition for their contributions, has raised concerns about the sustainability of the peer review system (21, 26). Addressing these systemic issues should occur in parallel with developing responsible policies for emerging technologies such as LLMs, to ensure that peer review continues to uphold the highest standards of scientific integrity.

Enhancing recognition for peer reviewers is an important element in improving the sustainability of the system. Platforms such as Publons, now integrated with Web of Science, allow reviewers to record, verify, and showcase their peer review contributions. Such systems offer a means of professional acknowledgment and may encourage participation in peer review. As LLM-assisted peer review becomes more common, clear attribution of reviewer effort, including the use and scope of LLM assistance, could help preserve the value of human judgment and promote transparency in reviewer contributions.

This article has several limitations. The proposed framework is based primarily on expert opinion and early observations, rather than empirical evaluation of LLM use in peer review settings. Moreover, the framework is tailored to the context of dental research and its generalizability to other disciplines has not been assessed. It is also important to acknowledge that the rapid advancement of machine learning poses a challenge for policy and guideline development. This underscores the need for continued monitoring and periodic refinement of guidance as machine learning tools continue to advance.

Future research should explore how LLMs are used in real-world peer review workflows, including the extent, frequency, and outcomes of their application. Empirical studies assessing the impact of LLM assistance on review quality, reviewer efficiency, and editorial decision-making will be essential (9). In addition, surveys and interviews with reviewers, editors, and authors could provide valuable insights into ethical perceptions, practical challenges, and best practices. Long-term, there is a need to examine whether and how responsible integration of LLMs can contribute to improving the peer review system in dental medicine.

Conclusions

LLMs entail both benefits and risks for dental peer review. When used responsibly, particularly in offline or locally hosted environments, LLMs can help reviewers improve the clarity and structure of their reports. However, without clear guidelines, transparency, and safeguards for confidentiality, their use could undermine trust, accountability, and scientific rigor. The framework presented here supports the responsible employment of LLMs while maintaining the fundamental standards of rigorous peer review in dental research publishing.

Acknowledgements

The authors deny any possible conflict of interest related to this work. The authors acknowledge the use of ChatGPT Plus (OpenAI, San Francisco, CA, USA) and DeepSeek R1 (High-Flyer, Hangzhou, Zhejiang, China) for language review, including grammar, spelling, punctuation, and linguistic refinement. The French summary was prepared with the assistance of DeepL (DeepL SE, Cologne, Germany) and ChatGPT Plus, based on the original German version.

Zusammenfassung

Wer hat das begutachtet? Ansätze für eine verantwortungsvolle Integration grosser Sprachmodelle in das Peer-Review zahnmedizinischer Fachartikel

Einleitung

Der Einsatz grosser Sprachmodelle (GSM) wie ChatGPT, Claude, DeepSeek oder Google Gemini nimmt in der Wissenschaft stetig zu – auch im Bereich der Zahnmedizin. Diese Modelle bieten Potenzial, die Qualität und Effizienz von Peer-Reviews zu verbessern, insbesondere bei Sprachbarrieren. Gleichzeitig wirft ihre Nutzung ethische und praktische Fragen auf, insbesondere in Bezug auf Vertraulichkeit, Verantwortlichkeit und Transparenz. Zwar können GSM zur sprachlichen Optimierung von Peer-Review-Berichten beitragen, doch mangelt es ihnen an der notwendigen fachspezifischen Expertise, um wissenschaftliche Inhalte sachgerecht zu bewerten. Ziel dieser Arbeit ist es, einen Überblick über die aktuellen Richtlinien führender zahnmedizinischer Fachzeitschriften zur GSM-Nutzung zu geben und ein praxisorientiertes Rahmenkonzept für deren verantwortungsvollen Einsatz im Peer-Review-Prozess in der Zahnmedizin zu entwickeln.

Material und Methoden

Im Zeitraum vom 22. bis 24. Juli 2025 wurden die Peer-Review-Richtlinien der 50 zahnmedizinischen Fachzeitschriften mit dem höchsten Journal Impact Factor (2024) systematisch analysiert. Die Untersuchung umfasste öffentlich zugängliche Informationen auf den Webseiten der Fachzeitschriften sowie Richtlinien der zugehörigen Verlage. Berücksichtigt wurden öffentlich zugängliche Informationen auf den Webseiten der Fachzeitschriften sowie ergänzende Vorgaben der jeweiligen Verlage. Untersucht wurden unter anderem der Peer-Review-Modus, Regelungen zur Vertraulichkeit, explizite Erwähnungen von GSM, erlaubte oder untersagte Nutzung, Offenlegungspflichten sowie Regelungen zur Verantwortlichkeit. Die Ergebnisse wurden deskriptiv ausgewertet und dienten als Grundlage für die Entwicklung eines Rahmenkonzepts.

Resultate

Von den 50 analysierten Fachzeitschriften erwähnten 43 (86 %) GSM oder künstliche Intelligenz explizit in ihren Richtlinien. 26 Fachzeitschriften (52 %) gestatteten den Einsatz von GSM unter bestimmten Voraussetzungen, 19 (38 %) untersagten die Nutzung vollständig, und 5 (10 %) enthielten keine klaren Regelungen. In den Fällen, in denen GSM zugelassen waren, war der Einsatz auf sprachliche Korrekturen und strukturelle Verbesserungen beschränkt – nicht jedoch auf inhaltliche oder wissenschaftliche Bewertungen. In 92 % der Fälle war bei erlaubtem Einsatz eine transparente Offenlegung der Nutzung verpflichtend. Die inhaltliche Verantwortung für die Begutachtung lag in allen Fällen weiterhin ausschliesslich beim menschlichen Reviewer. Die Analyse identifizierte zentrale Risiken wie potenzielle Datenschutzverletzungen, intransparente Autorschaft, fehlerhafte Inhalte, fehlende fachspezifische Kompetenz und eine unkontrollierte Nutzung ohne Offenlegung.

Diskussion

Die Ergebnisse verdeutlichen eine hohe Heterogenität in den Richtlinien zur GSM-Nutzung innerhalb zahnmedizinischer Fachzeitschriften. Daraus ergibt sich ein klarer Handlungsbedarf für die Entwicklung harmonisierter, regelmässig aktualisierter und disziplinspezifischer Leitlinien. Das vorgeschlagene Rahmenkonzept basiert auf fünf zentralen Prinzipien: (1) Reviewer müssen fachlich qualifiziert und persönlich für den Review-Inhalt verantwortlich bleiben; (2) Manuskripte dürfen nicht an online-basierte GSM übermittelt werden – nur lokal oder offline betriebene Modelle (z. B. GPT4ALL, Nomic AI, New York, NY, USA) sind datenschutzrechtlich weniger risikobehaftet; (3) jegliche Nutzung von GSM ist transparent offenzulegen; (4) der Einsatz darf sich ausschliesslich auf sprachliche und strukturelle Optimierung beschränken; (5) fachspezifische Schulungen sowie standardisierte Offenlegungsprozesse sollten etabliert werden. Diese Massnahmen können zur Wahrung der Integrität, wissenschaftlichen Strenge und Nachhaltigkeit des Peer-Review-Systems in der Zahnmedizin beitragen.

Résumé

Qui a évalué cela ? Approches pour une intégration responsable des grands modèles linguistiques dans la révision par les pairs d'articles dentaires spécialisés

Introduction

L'utilisation croissante des grands modèles de langage (LLM, Large Language Models) tels que ChatGPT, Claude, DeepSeek ou Google Gemini dans les milieux scientifiques, y compris en médecine dentaire, ouvre de nouvelles perspectives, notamment en matière d'amélioration de la qualité et de l'efficacité du processus de relecture par les pairs. Ces outils peuvent faciliter la rédaction, particulièrement lorsque des barrières linguistiques existent. Toutefois, leur intégration soulève des questions éthiques et pratiques majeures, notamment en ce qui concerne la confidentialité des données, la responsabilité individuelle du réviseur, et la transparence du processus. Bien que les LLM puissent contribuer à l'amélioration linguistique des rapports d'évaluation, ils ne disposent pas de l'expertise disciplinaire requise pour évaluer rigoureusement le contenu scientifique. Cette étude vise à examiner les politiques éditoriales actuelles des principales revues dentaires spécialisées concernant l'usage des LLM, et à proposer un cadre conceptuel pour leur intégration responsable dans l'évaluation scientifique.

Matériels et méthodes

Une analyse systématique a été réalisée entre le 22 et le 24 juillet 2025 sur les politiques éditoriales de 50 revues dentaires spécialisées, sélectionnées selon leur facteur d'impact (2024). Les données ont été extraites à partir des sites web officiels des revues, ainsi que des lignes directrices générales de leurs maisons d'édition. Les critères étudiés comprenaient : le modèle de relecture par les pairs, les exigences en matière de confidentialité, la mention explicite de l'intelligence artificielle ou des LLM, les autorisations ou interdictions d'usage, les obligations de divulgation, et les clauses de responsabilité. Les données ont été analysées de manière descriptive et ont servi de base à la formulation d'un cadre de recommandations.

Résultats

Parmi les 50 revues analysées, 43 (86 %) mentionnaient explicitement les LLM ou l'intelligence artificielle dans leurs politiques éditoriales. 26 revues (52 %) autorisaient l'utilisation des LLM sous conditions, 19 (38 %) l'interdisaient formellement, et 5 (10 %) ne formulaient aucune position claire. Lorsque l'usage était autorisé, il était strictement limité à des tâches non évaluatives telles que la correction linguistique ou l'amélioration structurelle du texte. Dans 92 % des cas d'autorisation, la divulgation explicite de l'usage des LLM était exigée. La responsabilité de l'évaluation restait, dans tous les cas, entièrement assumée par le réviseur humain. Les risques identifiés incluent notamment des atteintes à la confidentialité, un flou sur l'attribution de la paternité intellectuelle, des erreurs de contenu, un manque d'expertise disciplinaire, et un usage non déclaré des outils.

Discussion

Les résultats révèlent une hétérogénéité marquée dans les politiques éditoriales concernant l'usage des LLM dans les revues de médecine dentaire. Cela souligne la nécessité urgente de lignes directrices harmonisées, actualisées régulièrement et adaptées au contexte disciplinaire. Le cadre proposé repose sur cinq principes fondamentaux : (1) les évaluateurs doivent disposer de l'expertise nécessaire et demeurer responsables du contenu de leur rapport ; (2) les manuscrits ne doivent pas être introduits dans des interfaces LLM en ligne – seuls des modèles hors ligne ou localement hébergés (ex. : GPT4ALL, Nomic AI, New York, NY, USA) sont acceptables ; (3) toute utilisation d'un LLM doit faire l'objet d'une divulgation transparente ; (4) l'usage doit se limiter à l'optimisation linguistique ou structurelle, sans évaluation du contenu scientifique ; (5) des formations spécifiques et des procédures normalisées de divulgation doivent être mises en œuvre. Ces mesures peuvent contribuer à préserver l'intégrité, la rigueur scientifique et la durabilité du système d'évaluation par les pairs en médecine dentaire.

References

1. Eggmann F, Weiger R, Zitzmann NU, Blatz MB. Implications of large language models such as chatgpt for dental medicine. J Esthet Restor Dent. 2023;35(7):1098-102.

2. Bornstein MM, Staedler YM, Cattin PC. Artificial intelligence will change the research environment in dental medicine dramatically: Will algorithms replace literature reviews in the near future? Dentomaxillofac Radiol. 2025;54(5):329-31.

3. Büttner M, Leser U, Schneider L, Schwendicke F. Natural language processing: Chances and challenges in dentistry. J Dent. 2024;141:104796.

4. Fuchs A, Trachsel T, Weiger R, Eggmann F. ChatGPT's performance in dentistry and allergyimmunology assessments: A comparative study. Swiss Dent J. 2023;134(2):1-17.

5. Uribe SE, Maldupa I. Estimating the use of ChatGPT in dental research publications. J Dent. 2024;149:105275.

6. Rokhshad R, Bagherianlemraski M, Ehsani SS, Haghighat S, Schwendicke F. Large language models for the screening step in systematic reviews in dentistry. J Dent. 2025;160:105877.

7. Biswas S, Dobaria D, Cohen HL. ChatGPT and the future of journal reviews: A feasibility study. Yale J Biol Med. 2023;96(3):415-20.

8. Saad A, Jenko N, Ariyaratne S, Birch N, Iyengar KP, Davies AM, et al. Exploring the potential of ChatGPT in the peer review process: An observational study. Diabetes Metab Syndr. 2024;18(2):102946.

9. Marrella D, Jiang S, Ipaktchi K, Liverneaux P. Comparing AI-generated and human peer reviews: A study on 11 articles. Hand Surg Rehabil. 2025:102225.

10. Checco A, Bracciale L, Loreti P, Pinfield S, Bianchi G. AI-assisted peer review. Humanities and social sciences communications. 2021;8(1):1-11.

11. Lo Vecchio N. Personal experience with AI-generated peer reviews: A case study. Research Integrity and Peer Review. 2025;10(1):4.

12. Schintler LA, McNeely CL, Witte J. A critical examination of the ethics of AI-mediated peer review. arXiv preprint arXiv:230912356. 2023.

13. Hosseini M, Horbach SP. Fighting reviewer fatigue or amplifying bias? Considerations and recommendations for use of chatgpt and other large language models in scholarly peer review. Research Integrity and Peer Review. 2023;8(1):4.

14. BaHammam AS. Peer review in the artificial intelligence era: A call for developing responsible integration guidelines. Nat Sci Sleep. 2025;17:159-64.

15. Joachim MV, Dodson TB, Laviv A. How artificial intelligence differs from humans in peer review. J Oral Maxillofac Surg. 2025;83(8):1040-50.

16. Li ZQ, Xu HL, Cao HJ, Liu ZL, Fei YT, Liu JP. Use of artificial intelligence in peer review among top 100 medical journals. JAMA Netw Open. 2024;7(12):e2448609.

17. Suchikova Y, Tsybuliak N, Teixeira da Silva JA, Nazarovets S. Gaidet (generative AI delegation taxonomy): A taxonomy for humans to delegate tasks to generative artificial intelligence in scientific research and publishing. Account Res. 2025:1-27. doi: 10.1080/08989621.2025.2544331.

18. Wiwanitmkit S, Wiwanitkit V. Artificial intelligence, academic publishing, scientific writing, peer review, and ethics. Braz J Cardiovasc Surg. 2024;39(4):e20230377.

19. Liang W, Izzo Z, Zhang Y, Lepp H, Cao H, Zhao X, et al. Monitoring AI-modified content at scale: A case study on the impact of ChatGPT on AI conference peer reviews. In: Ruslan S, Zico K, Katherine H, Adrian W, Nuria O, Jonathan S, et al., editors. Proceedings of the 41st International Conference on Machine Learning; Proceedings of Machine Learning Research: PMLR; 2024. p. 29575--620.

20. Poisot T. Here is a revised version of your review with improved clarity [Internet]. https://buttondown.com/ctrl-alt-tim/archive/vol-24-here-is-a-revised-version-of-your-review/ (last accessed on August 19, 2025). 2025. [cited 2025].

21. Routledge D, Pariente N. On improving the sustainability of peer review. PLoS Biol. 2025;23(3):e3003127.

22. Weber MT, Noll R, Marchl A, Facchinello C, Grünewaldt A, Hügel C, et al. Medbot vs realdoc: Efficacy of large language modeling in physician-patient communication for rare diseases. J Am Med Inform Assoc. 2025;32(5):775-83.

23. Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Intern Med. 2023;183(6):589-96.

24. Lin Z. Hidden prompts in manuscripts exploit AI-assisted peer review. arXiv preprint arXiv:250706185. 2025.

25. Rich AS, Gureckis TM. Lessons for artificial intelligence from the study of natural stupidity. Nature Machine Intelligence. 2019;1(4):174-80.

26. Gruda D. Three AI-powered steps to faster, smarter peer review. Nature. 2025. doi: 10.1038/d41586-025-00526-0.

27. STM. Recommendations for a classification of AI use in academic manuscript preparation [Internet]. https://stm-assoc.org/document/recommendations-for-a-classification-of-ai-use-in-academic-manuscript-preparation/2025 (last accessed on August 19, 2025). 2025. [cited 2025].