Recommendations for disclosure of artificial intelligence in scientific writing and publishing: a regional anesthesia and pain medicine modified Delphi study

FETTIPLACE, Michael R <http://orcid.org/0000-0001-6946-8355>, BHATIA, Anuj <http://orcid.org/0000-0001-6292-8654>, CHEN, Yian <http://orcid.org/0000-0002-5971-9111>, OREBAUGH, Steven L <http://orcid.org/0000-0001-9517-6575>, GOFELD, Michael <http://orcid.org/0000-0001-7372-4171>, GABRIEL, Rodney A <http://orcid.org/0000-0003-4443-0021>, SESSLER, Daniel I <http://orcid.org/0000-0001-9932-3077>, LONSDALE, Hannah <http://orcid.org/0000-0001-9953-8844>, BUNGART, Brittani <http://orcid.org/0000-0001-5705-3742>, CHENG, Christopher P <http://orcid.org/0000-0001-9112-9464>, BURNETT, Garrett W <http://orcid.org/0000-0003-0565-9499>, HAN, Lichy, WILES, Matthew <http://orcid.org/0000-0003-2299-0060>, COPPENS, Steve <http://orcid.org/0000-0002-6026-7757>, JOSEPH, Thomas, SCHREIBER, Kristin L <http://orcid.org/0000-0002-7361-2214>, VOLK, Thomas <http://orcid.org/0000-0001-5175-7159>, URMAN, Richard D, KOVACHEVA, Vesela P, WU, Christopher L <http://orcid.org/0000-0002-4484-0787>, MARIANO, Edward R <http://orcid.org/0000-0003-2735-248X> and IP, Vivian H Y <http://orcid.org/0000-0001-6158-4415>

OPEN ACCESS

Recommendations for disclosure of artificial intelligence in scientific writing and publishing: a regional anesthesia and pain medicine modified Delphi study

Michael R Fettiplace ,[1] Anuj Bhatia ,[2,3] Yian Chen ,[4] Steven L Orebaugh ,[5] Michael Gofeld ,[6] Rodney A Gabriel ,[7] Daniel I Sessler ,[8,9] Hannah Lonsdale ,[10] Brittani Bungart ,[10] Christopher P Cheng ,[11] Garrett W Burnett ,[11] Lichy Han,[12] Matthew Wiles ,[13,14] Steve Coppens ,[15,16] Thomas Joseph,[17] Kristin L Schreiber ,[18] Thomas Volk ,[19,20] Richard D Urman,[21] Vesela P Kovacheva,[18] Christopher L Wu ,[22,23] Edward R Mariano ,[12,24] Vivian H Y Ip [25]

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/rapm-2025-106852).

For numbered affiliations see end of article.

Correspondence to
Dr Michael R Fettiplace; fettiplace@gmail.com

Received 19 May 2025
Accepted 17 June 2025

To cite: Fettiplace MR, Bhatia A, Chen Y, *et al*. *Reg Anesth Pain Med* Epub ahead of print: [*please include* Day Month Year]. doi:10.1136/rapm-2025-106852

ABSTRACT

Introduction The use of artificial intelligence (AI) in the scientific process is advancing at a remarkable speed, thanks to continued innovations in large language models. While AI provides widespread benefits, including editing for fluency and clarity, it also has drawbacks, including fabricated content, perpetuation of bias, and lack of accountability. The editorial board of *Regional Anesthesia & Pain Medicine* (RAPM) therefore sought to develop best practices for AI usage and disclosure.

Methods A steering committee from the American Society of Regional Anesthesia and Pain Medicine used a modified Delphi process to address definitions, disclosure requirements, authorship standards, and editorial oversight for AI use in publishing. The committee reviewed existing publication guidelines and identified areas of ambiguity, which were translated into questions and distributed to an expert workgroup of authors, reviewers, editors, and AI researchers.

Results Two survey rounds, with 91% and 87% response rates, were followed by focused discussion and clarification to identify consensus recommendations. The workgroup achieved consensus on recommendations to authors about definitions of AI, required items to report, disclosure locations, authorship stipulations, and AI use during manuscript preparation. The workgroup formulated recommendations to reviewers about monitoring and evaluating the responsible use of AI in the review process, including the endorsement of AI-detection software, identification of concerns about undisclosed AI use, situations where AI use may necessitate the rejection of a manuscript, and use of checklists in the review process. Finally, there was consensus about AI-driven work, including required and optional disclosures and the use of checklists for AI-associated research.

Discussion Our modified Delphi study identified practical recommendations on AI use during the scientific writing and editorial process. The workgroup highlighted the need for transparency, human accountability, protection of patient confidentiality, editorial oversight, and the need for iterative updates. The proposed framework enables authors and editors to harness AI's efficiencies while maintaining the fundamental principles of responsible scientific communication and may serve as an example for other journals.

INTRODUCTION

Artificial intelligence (AI), with a particular emphasis on generative AI, has advanced at a remarkable pace, ushering in large-language models (LLMs) such as ChatGPT (OpenAI, San Francisco, CA),[1] Claude (Anthropic Inc, San Francisco, CA),[2] and Gemini (Google, Mountain View, CA)[3] that can interpret and generate text with striking fluency.[4 5] AI tools can expedite literature searches, refine manuscript drafts, and even assist with data analysis.[6] However, the use of AI raises substantial issues for journal editorial boards, including the board of *Regional Anesthesia and Pain Medicine (RAPM)*. AI stands to save researchers valuable time, reduce costs, and improve writing quality—particularly for writers whose primary language is not English and for busy clinicians. Conversely, unchecked use of LLMs carries the risk of introducing fabricated references, stifling original ideas, perpetuating biases from flawed training data, and diluting accountability if human authors rely too heavily on automated text generation without verification.[7 8]

Guidelines exist for reporting on AI-centered research methodologies (eg, MINIMAR, CONSORT-AI, DECIDE-AI).[9–11] Yet, much of AI use happens "behind the scenes" with grammar assistance software like Grammarly or Microsoft Copilot, which authors use frequently but may not recognize as AI. The proliferation of these tools raises concerns about how best to track AI involvement so that published research remains trustworthy, transparent, and reproducible. Moreover, sophisticated LLMs can now create entire review articles that appear plausible yet may be riddled

with "hallucinated" information—fabricated data, stereotyped images, non-existent references, or even biased language, thereby threatening the integrity of scientific discourse.[12 13] Various aspects of AI thus contribute to challenge journal editors with the competing demands of encouraging use of AI tools to improve the clarity and fluency of articles while limiting fraudulent submissions.[14 15]

Given these concerns, the editorial board of *RAPM* was tasked to identify recommendations about using and disclosing AI for articles submitted to *RAPM*. A steering committee of members of the American Society of Regional Anesthesia and Pain Medicine (ASRA) screened the literature for AI-associated issues in publishing along with current standards for reporting use of AI. The steering committee identified heterogeneity of requirements across journals, including unclear definitions and discrepancies about what constitutes AI, inadequate details about appropriate disclosure of AI use, unclear guidance to authors about the use of AI, lack of ground rules about AI in the review process, and divergent recommendations around checklists or tools to standardize the disclosure and evaluation of AI use (online supplemental table 1). Given these discrepancies, the steering committee developed a set of key questions aimed at establishing consensus on use and disclosure of AI in scientific writing. We then conducted a modified Delphi consensus study[16] with an expert workgroup to identify recommendations for authors, reviewers, and AI-driven[17] research for *RAPM*.

METHODS

Composition of the steering committee and workgroup

This *RAPM*-led Delphi consensus study was approved by the *RAPM* Editor-in-Chief (EIC). A steering committee was comprised of members of the *RAPM* editorial board who chose to participate. The steering committee was responsible for background research, question generation (online supplemental table 2), and participant identification. For the larger workgroup, the steering committee identified members of the *ASRA Pain Medicine* community, who were experienced in the use of AI tools, along with members of the anesthesia community who had previously published on AI topics in anesthesia journals, EICs of anesthesia journals, and EICs of prominent medical journals. An additional solicitation was disseminated via X (formerly Twitter) to the *ASRA Pain Medicine* community members for participation. Potential participants were queried via a web form about their desire to participate, their experiences with AI, research, medicine, editorial work, and their experience publishing research in medicine-related fields, including anesthesiology. If an invitee did not feel qualified, they were encouraged to identify an alternative with sufficient experience and time to contribute.

Modified Delphi process

The enclosed study was granted a waiver from the University of Illinois at Chicago institutional review board given the survey methodology and minimal risk to participants. Based on responses in the preliminary screening, questions were generated for the workgroup (online supplemental table 1). Following question generation, a modified Delphi methodology was applied to identify consensus. There were two rounds of electronic voting followed by a third round of commentary. Contributors' responses during the first two rounds were kept anonymous to ensure that each collaborator's opinion was equally weighted. Following the first round, additional questions and commentary were added. A second round of re-voting was conducted for new questions with clarification on old questions. All questions were included in the second round with median score, IQR and comments from the first round. A third round (non-voting) was conducted to provide clarifications on points of disagreement. Those unable to attend the meetings were provided an opportunity to respond via a Google form with comments. At least two follow-up reminders were provided to encourage participation for all rounds. Numerical answers from the second round were evaluated for disagreement using the Rand/University of California at Los Angeles (UCLA) process.[18] In brief, this involved the calculation of a disagreement index (among participants) using interpercentile ranges. Answers with support or oppose rating and low disagreement indices were included as recommendations. Full details of the consensus process are described in the supplemental methods.

Artificial intelligence statement

ChatGPT 4.0 (OpenAI, San Francisco CA, no customizations) was used in the ideation and research for this article (July 2024). ChatGPT o1 (OpenAI, San Francisco CA, February and March 2025) was used for help in drafting the introduction and discussion. Grammarly was used for help in grammar and fluency (Grammarly, San Francisco, CA; February 2025).

RESULTS

Seventy-one participants were invited with a total of forty-five agreeing to participate. Forty-one (91%) participants responded in the first round, 39 (87%) in the second round, and 30 (67%) participated in clarification sessions (online supplemental figure 1). Participants were invited on 6 October 2024, with replies required by 22 October 2024. On 23 October 2024, the first round of voting was distributed, with responses required by November 11. The second round was distributed on 19 November 2024, with replies required by December 8. The third round of voting was conducted between December 2024 and January 2025. For details regarding participants who completed at least one round of voting, with accompanying qualifications, please see online supplemental table 3).

Part 1: recommendations to Authors

The workgroup was initially asked what should qualify as AI with items listed in box 1 that reached a strong consensus. There was broad agreement from the workgroup that we should be inclusive of what constitutes AI given the rapid progress in the field. Next, the workgroup was asked what should be disclosed during the submission process, with points in box 2 reaching strong consensus. There was a desire to ensure that work

Box 1 Items defined as artificial intelligence (AI) by the workgroup

⇒ Large-language models (VSA).
⇒ Generative AI for image or figure generation (VSA).
⇒ Computer vision and image analysis including interpretation of graphs or figures (VSA).
⇒ Machine learning algorithms and tools (SA).
⇒ Natural language processing (NLP) and NLP tools (SA).
⇒ Chatbots (SA).
⇒ Robotic tools with an internal decision-making system (SA).
⇒ Augmented search tools (A).
⇒ Predictive and data analytics methods (A).

(A) agree; (SA) strongly agree; (VSA) very strongly agree

Box 2 Disclosure details in the submission process

2.1 What needs to be disclosed

⇒ Artificial intelligence (AI) systems used as methods should include *software package,* the *specific model*, *version*, *publisher*, *location of publisher,* and *date used* (VSA).
⇒ Any possible non-transparent or non-intuitive components introduced by AI utilization (VSA).
⇒ Writing help, including drafting and editing (VSA).
⇒ Drafting a manuscript from scratch (VSA).
⇒ Hypothesis generation (VSA).
⇒ Background research (eg, identification of gaps in the literature by an large-language model (LLM)) (VSA).
⇒ Data analysis (VSA).
⇒ Interpretation of data, tables, and figures (eg, use of LLM image analysis) (VSA).
⇒ Methods development (VSA).
⇒ Audio or video generation (and/or interpretation of audio input/visual input) (VSA).
⇒ Ethical concerns of AI use (VSA).
⇒ Perpetuation of systemic bias by AI (VSA).
⇒ Translation of manuscript from another language (SA).
⇒ Table and figure generation (SA).
⇒ Reference identification (SA).

2.2 Where the disclosure should be made

⇒ Online submission system (VSA).
⇒ Methods section (VSA).
⇒ Cover letter (SA).

(A) agree; (SA) strongly agree; (VSA) very strongly agree

presented in a manuscript be replicable, and thus the workgroup asserted that authors should report sufficient details for replication and identification of associated bias and errors introduced by AI software.

The workgroup also agreed that authors should identify any non-transparent or non-intuitive components introduced by AI use which might arise from hidden layers of logic in an LLM or ML algorithm, which may influence the reported outcomes. For example, neural networks and LLMs often contain hidden layers of reasoning, whereby an AI system can generate fabricated information (eg, references), without a logical source; therefore, authors should disclose AI systems that include hidden layers of logic. The authors should also identify any possible ethical concerns introduced by AI, as has been discussed elsewhere.[19 20] Examples of ethical concerns include inputting protected health information or unpublished data into AI systems, thus violating privacy rights or degrading scientific ownership rights. Disclosures of AI use should be made both in the manuscript and online submission system (box 2). A few items related to definitions and disclosures did not reach consensus and/or were neutral in recommendation as described in the supplemental results (online supplemental boxes 1 and 2). As a specific supplemental point, the workgroup asserted that reporting the grammatical rewriting of a single sentence by Microsoft Copilot in Word (or similar) was unnecessary, but more extensive rewriting may merit disclosure.

Regarding authorship, the workgroup agreed that AI cannot be an author, and human authors must take responsibility for the work (box 3). A corollary is that human authors must critically review and verify AI-assisted work and attest to this verification process. If an AI is the only member of the authorship team

Box 3 Recommendations to authors about manuscript preparation

3.1 Authorship stipulations

⇒ AI cannot be listed as an author (VSA).
⇒ If an large-language model (LLM) met International Committee of Medical Journal Editors (ICMJE) authorship criteria, it cannot be an author (VSA).
⇒ If an LLM is the only member of the team that meets ICMJE criteria, it cannot be an author (VSA).
⇒ LLM produced solutions require validation or corroboration by a human team prior to consideration for publication (VSA).
⇒ Human authors should take final responsibility for the manuscript (VSA).
⇒ Authors should double-check AI-assisted work (VSA).
⇒ Authors should attest to checking AI work and taking responsibility (VSA).
⇒ Articles with significant AI involvement should include a team member with AI expertise (SA).

3.2 Recommendations about AI use

⇒ Authors should not input their manuscript into an LLM if there are active ethical concerns about an LLM (VSA).
⇒ Authors should provide a searchable index for all referenced citations (VSA).
⇒ Authors can use AI exclusively to improve language and readability (SA).
⇒ Manuscripts can be input into LLMs for editing purposes as long as ethical and privacy concerns are addressed (A).
⇒ The authors can input parts (eg, sections or paragraphs) of their manuscript into an LLM if data is deidentified (A).

3.3 Recommendations against submission

⇒ If authors do not understand the translated English, word choice, or grammar (SA).
⇒ If authors do not understand how the data was analyzed by the AI (VSA).
⇒ If authors do not understand what an AI-generated figure means (VSA).
⇒ If authors do not understand the hypothesis generation (VSA).
⇒ If the authors have not verified the citations (VSA).
⇒ If there are any conclusions from AI that the authors cannot explain or justify (VSA).
⇒ Review articles that are purely drafted and edited by an LLM (VSA).
⇒ If a significant part (including the entirety) of the discussion was drafted and edited by an LLM (VSA).
⇒ If a manuscript has been input into an LLM with active concern for plagiarism (VSA).
⇒ If a manuscript has been input into an LLM with active concern for privacy violations or health insurance portability and accountability act (HIPAA) violations (VSA).

3.4 Additional requests/stipulations

⇒ The journal can request submission of datasets (VSA).
⇒ The journal can request screenshots or other methods of validation of AI usage including a model reference or a copy of the conversation (VSA).
⇒ For custom data analysis, the journal should request the code used to perform statistical analysis (SA).

Continued

Box 3 Continued

⇒ There should be a journal ban or other repercussions for authors who do not properly disclose AI usage as outlined in the recommendations (A).

(A) agree; (SA) strongly agree; (VSA) very strongly agree

that meets ICJME criteria, we would not view this as a valid submission. Finally, if an article involves significant AI use, the workgroup recommends the involvement of an author(s) with AI expertise similar to the involvement of a statistician in clinical study analyses and a librarian in systematic reviews. Significant AI use is at the discretion of the authors and reviewers.

Recommendations about AI use in the writing process are found in box 3. The workgroup agreed that authors can input data into LLMs if privacy and ethical standards are maintained. There was strong consensus from the workgroup that authors can use AI exclusively to improve the language and readability via translation services and grammar services without consulting a native or fluent language speaker. This may involve entering the entire manuscript into an LLM system for analysis. The workgroup agreed that language improvement is a substantial benefit that AI systems provide to the scientific community.

When authors are preparing to submit a manuscript, the workgroup agreed that comprehension of the hypotheses, methodology, wording, detailed process of data analysis, figures, and diagrams is crucial. Under certain circumstances, the journal should caution authors against submission of their manuscript if they do not comprehend the work performed by and attributed to the AI system, with the authors using their own discretion. The workgroup recommended against submission of a manuscript, in certain circumstances, listed in box 3. As a specific point, the workgroup recommends against submission of review articles (including narrative reviews, scoping reviews, systematic reviews, and other reviews) that are largely drafted and edited by an LLM.

The workgroup felt strongly that the journal should be able to request additional information including datasets or screenshots (box 3). Consequences for non-compliance with requests for information may impact present and future manuscript submissions to RAPM (ie, including prohibition from submitting manuscripts in the future). Several other items related to manuscript preparation did not reach consensus and/or were neutral in recommendation as described in the supplemental results (online supplemental box 3).

Part 2: recommendations to editors/reviewers

The workgroup also considered the use of AI in the review process itself. The workgroup was queried about the option for the EIC and/or reviewers to use automated software to detect AI use in manuscript preparation. There was consensus that the journal should use software to detect AI use in submitted manuscripts (box 4). The workgroup also reached consensus that reviewers could use AI in the review process but with certain limitations. As a specific point, the workgroup agreed that the journal should use AI detection software but also agreed that reviewers should not put manuscript contents into AI tools (including AI detection software) or automatically use AI tools to evaluate the manuscript. Reviewers should coordinate with the EIC and senior editors on these topics. Reviewers can use AI tools like OpenAI's ChatGPT,[1] Google's Gemini,[3] or Open Evidence[21] to summarize scientific topics related to manuscripts but should not ask a chatbot to review a manuscript's content.

In the case of undisclosed AI use, the workgroup recommended that any member of the review and reader process can identify a concern for undisclosed AI use. This concern should trigger a critical review of the manuscript for undisclosed AI use (box 4). Beyond undisclosed AI use, the workgroup recommended that in certain situations, a manuscript should be rejected from consideration (box 4), primarily in the case of ethical violations and plagiarism. A few items about use of AI in the review process did not reach consensus and/or were neutral in recommendation as

Box 4 The evaluation of artificial intelligence (AI) use for reviewers and editors

4.1 Use of AI software in the review process

⇒ Metadata can be checked to identify concerns about fabricated data (VSA).
⇒ The journal should use software to detect AI use in a submitted manuscript (SA).
⇒ Reviewers should not automatically use AI tools to review the manuscript (SA).
⇒ In the case that editors/reviewers use AI, every reviewer should be required to respond to a question about whether they used AI in the review process (SA).
⇒ In the case that editors/reviewers use AI, every reviewer should be required to disclose all details of AI utilization (SA).
⇒ Reviewers can use AI tools to review the manuscript as long as none of the original manuscript content (eg, figures, tables, images, or text) are input into an AI tool (*eg, a conversation with an AI tool about the content*) (A).
⇒ The editor/EIC can use AI tools to review the manuscript in any form (A).
⇒ The editor/EIC can use AI tools to review the manuscript, if none of the original manuscript content (eg, figures, tables, images, or text) is input into an AI tool (A).

4.2 Who can identify concerns about undisclosed AI use in manuscripts

⇒ The Editor-in-Chief (SA).
⇒ An editorial board member (SA).
⇒ A reviewer (SA).
⇒ The editorial coordinator (SA).
⇒ The publisher (SA).
⇒ Readers after publication (A).

4.3 Rejection stipulations

⇒ Concern about lack of transparency or lack of disclosure about AI use (VSA).
⇒ The editorial team can provide an opportunity for the authors to provide clarification in the case of lack of transparency (VSA).
⇒ Violation of restrictions and cautions as stipulated in "recommendations to authors" (A).
⇒ Concern from editors/reviewers about AI-driven plagiarism (A).
⇒ Concern from editors/reviewers about AI-fabricated data or AI-fabricated analysis (A).
⇒ Concern from editors/reviewers about ethical violations by AI (A).

(A) agree; (SA) strongly agree; (VSA) very strongly agree

Box 5 Use of checklists in the review process

5.1 What manuscripts should require a checklist for artificial intelligence (AI) use

⇒ The journal should use an AI checklist for all papers (A).

5.2 Recommended checklist items

⇒ The authors should confirm that humans checked all AI-associated work for accuracy (VSA).
⇒ Authors should confirm that they comprehend and are responsible for all AI-associated conclusions (VSA).
⇒ Authors should confirm they are fully accountable for the work (VSA).
⇒ Authors should assert that they understand there may be repercussions for failing to disclose AI use (VSA).
⇒ Authors should check AI-generated figures for errors or misinformation (VSA).
⇒ Authors should check AI-generated tables for errors or misinformation (VSA).
⇒ Authors should confirm that AI use is appropriately identified for all the components where AI was used in their research (VSA).
⇒ Authors should attest to verifying cited references (given concerns about reference fabrication/hallucination) (SA).
⇒ Authors should check text for errors and misinformation that may have been introduced by an large-language model (SA).
⇒ Authors should confirm how AI-associated bias was accounted for (SA).
⇒ Authors should confirm how and where they protected individual privacy while using AI (SA).
⇒ There should be a free text section to detail the AI use (SA).
⇒ Authors should confirm that sufficient details are provided so that AI-driven methodologies can be replicated (A).

(A) agree; (SA) strongly agree; (VSA) very strongly agree

described in the supplemental results (online supplemental boxes 4 and 5).

Next, the workgroup considered the use of checklists to standardize the evaluation process of AI use in manuscripts submitted to RAPM. There was consensus that the journal should use a checklist for all submitted manuscripts (box 5). Items that reached consensus for inclusion in the checklist are listed in box 5.

Part 3: Specific recommendations for artificial intelligence (AI)-driven research

Finally, the workgroup considered a subset of recommendations that were specific to AI-driven research or interventions. AI-driven research is often independent of AI-assisted research. While AI-assisted research may be aimed at improving the writing quality of a manuscript, AI-driven research will investigate AI interventions. As examples, this includes the use of ML tools for anatomical structure identification as an intervention to accelerate ultrasound-guided nerve blocks or for identification and quantification of perioperative risk.[22] For AI-driven research, the workgroup agreed that the editorial staff could request additional items for submissions (box 6). Again, these recommendations were influenced by a desire for transparency and replicability in the research process.

For AI-specific interventions, there was also a recommendation to use AI-associated checklists. The workgroup agreed that

Box 6 Recommendations for artificial intelligence (AI)-driven research

6.1 In the case of custom AI, the following can be required for submission

⇒ Raw data (annotated and unannotated) used to train the AI or machine learning (ML) algorithm/code (SA).
⇒ Trained AI or ML algorithm/model (SA).
⇒ Code used to perform any data augmentation or normalization (SA).
⇒ The AI or ML algorithm/code (A).
⇒ Un-trained AI or ML algorithm/model (A).
⇒ Deposition of the trained AI or ML algorithm/model into the public domain (A).

6.2 Required checklists for AI-driven research

⇒ The journal should mandate a standardized checklist for AI-driven publications (VSA).
⇒ The appropriate checklist can be judged based on reviewer/editor input (SA).

(A) agree; (SA) strongly agree; (VSA) very strongly agree

the appropriate checklist can be judged based on reviewer/editor input and the choice of checklist may include the EQUATOR network recommendations, CONSORT-AI,[11] DECIDE-AI,[10] MINIMAR,[9] or other checklists. A few items related to AI-driven research did not reach consensus and/or were neutral in recommendation as described in the supplemental results (online supplemental box 6).

DISCUSSION

Our modified Delphi consensus study serves as a guide—for authors, reviewers, and editors—with practical recommendations on the use of AI in the editorial process, but may also inform peer-reviewers and readers of scientific manuscripts. While AI tools can accelerate research output, persistent challenges remain, including "hallucinated" references, biased training data, plagiarism, and diminished authorship accountability. In reviewing the available literature,[7 8 23–25] we found that definitions of AI often remain vague or are limited to a single tool (eg, ChatGPT). Many high-impact journals, such as *Science*, *Nature*, *The Lancet Digital Health*, and *Radiology*, have published commentaries urging more robust oversight of AI use in scholarly communication,[4 14 26–28] but few have implemented comprehensive policies. The most complete recommendations we found were provided by JAMA,[15] CELL,[29] and the Journal of Clinical Anesthesia,[14] which lay out editorial guidance on appropriate use (online supplemental table 1). Other journals, including the *New England Journal of Medicine*, defer to ICMJE recommendations.[30] Finally, The Lancet maintains a perspective that is at odds with the rapid adoption of AI in research,[31 32] asserting that AI may only be used to improve readability rather than to replace key researcher tasks.[33] The recommendations of the workgroup are largely consistent with prior guidance but arguably extend and clarify previous recommendations as summarized below.

Summary of recommendations

Part 1: Recommendations for authors

The workgroup adopted a broad definition of AI (encompassing LLMs, ML, and natural language processing tools) to reflect

the rapid pace of development. In fields like search (eg, Google Gemini) and word processing (Microsoft Copilot), LLM-based tools are now supplanting older approaches. To enhance transparency, authors should disclose any AI use—detailing the specific tools, version/model, publisher, and date of use—so others can replicate their work. The intent of these disclosures is transparency and replicability, two of the fundamental requirements of scientific research.[34] The workgroup advised that the disclosure should appear both in the online submission system (for editorial awareness) and in the Methods section (for readers). Authors should also describe possible ethical or methodological issues introduced by these tools, such as cognitive and automation bias, data privacy, transparency, hallucinations, misinformation, and lack of accountability and/or responsibility by the LLM.[20 35–39]

Consistent with other guidelines,[14 15 33] human authors bear final responsibility for all AI-assisted text, data, or analysis. AI tools cannot be listed as authors, and humans must validate the accuracy of AI-generated content.[40] While editing tools like Grammarly[41] can help refine language, the workgroup opposed fully AI-drafted articles. With the recent advent of OpenAI's o3 Deep Research, there have been assertions that AI can be the sole manuscript author,[40] but the workgroup was explicitly opposed to this. All co-authors must understand any AI-assisted contributions, verifying and clarifying them before submission.

Patient data privacy remains paramount.[42] Authors should deidentify patient information prior to using commercial or publicly available AI-based services and note that some sophisticated algorithms may be able to re-identify data. If an AI tool introduces privacy or ethical violations, the manuscript may be subject to rejection. The workgroup supports requiring raw data and LLM conversation transcripts for verification, aligning with US National Institutes of Health policies on data availability.[43] Recognizing the complexities of data sharing,[44 45] the workgroup also endorses limited-access repositories that simultaneously safeguard sensitive data and promote transparency. Failure to disclose AI use is deemed an ethical violation by ICMJE[30] and may result in rejection or retraction of a manuscript. For many points, we defer to local Institutional Review Boards or ethics panels that are charged with overseeing data privacy compliance.[46] However, if the editorial staff or reviewers identify lapses in ethical conduct or patient privacy and confidentiality during the review process, articles may be subject to rejection. The workgroup recommended that authors who engage in AI-assisted misconduct should be subject to consequences similar to other situations of research misconduct.

Part 2: Recommendations for the editorial and review process

The workgroup agreed that editors and reviewers could use AI tools cautiously, especially for language clarity or editing of reviews, but should not upload confidential manuscript content to tools that retain inputted data. This recommendation mirrors the US National Institutes of Health and National Science Foundation guidance, which prohibits uploading or sharing grant-related content with generative AI due to confidentiality. AI use by reviewers should be disclosed to the EIC. These recommendations are consistent with recommendations from COPE,[47] WAME,[48] and JAMA,[15 49] that emphasize transparency and appropriate disclosures.

The workgroup also agreed that *RAPM* should use AI-detection software to identify undisclosed AI-generated text, with the understanding that these tools have limitations (eg, false positives for human-written text).[50] Clear violations—such as plagiarism, falsified data, or undisclosed AI use—warrant rejection. Ultimately, human authors remain responsible for the manuscript's originality and integrity. The workgroup supports using standardized checklists at submission to confirm and clarify the extent of AI involvement.

Part 3: Recommendations for artificial intelligence (AI)-specific papers

For studies using AI as an intervention or methodological component (ie, AI-driven), the workgroup recommends permitting editors and reviewers to request relevant code and datasets. This ensures replicability and quality assessment. Authors may apply existing checklists (eg, MINIMAR,[9] CONSORT-AI,[11] DECIDE-AI[10]) to guide reporting. The journal can choose which AI-reporting guidelines best fit each submission, promoting flexibility without sacrificing rigor.

Key takeaways

The findings reported here parallel or expand on the broader discourse in the literature[14 15] and reflect the *ICMJE* requirement that only humans can be held to ethical and legal standards of authorship.[30] Collectively, the Delphi findings signal a path forward for *RAPM* with a few critical components:

1. Transparency: Authors should fully disclose how AI was used, whether for grammar checks or data analysis, in accordance with accepted journal policies.
2. Human accountability: Only human authors can bear ultimate responsibility for the content that mirrors international guidelines[30]
3. Privacy and confidentiality: Upholding patient data protection is non-negotiable, consistent with ethical approaches that have been the standard of human research for decades.[51]
4. *Editorial oversight:* AI-detection tools, requests for code or datasets, and flexible editorial policies can help ensure integrity without overburdening authors of valid submissions.
5. *Periodic updates:* Given AI's rapid evolution, these guidelines will require regular re-evaluation to remain current.

Limitations

As with any Delphi study, limitations stem from participant variability, personal biases, and iterative feedback. The participants were primarily researchers and clinicians in anesthesiology, and the workgroup did not include experts outside of medicine or experts with primary expertise in ethics. Additionally, we provided an extensive question list, but it was not exhaustive, and we likely missed subpoints. These results are also consensus recommendations, but implementation of policy requires expert judgment from the EIC. Ongoing rapid technological advances mean these recommendations may need updating sooner than traditional guidelines. The workgroup, therefore, advises a 2-year review cycle but will update sooner if prompted by significant changes within the field.

CONCLUSION

The rapid development of novel AI tools means that many authors will use them—intentionally or otherwise—throughout the research process from concept development to manuscript preparation. To maintain accuracy, transparency, and replicability, the workgroup advises that *RAPM* should adopt clear protocols to ensure transparency and replicability, consistent with good tenets of scientific research.[34] These protocols include recommendations about full disclosure of AI use, acceptance of responsibility by human authors, proper data availability, and robust reporting of AI's contributions and limitations. These

steps align with the broader goal of preserving scientific integrity and patient confidentiality, while leveraging AI's potential benefits for speed and efficiency. A balanced approach—welcoming innovation yet enforcing high ethical standards—will safeguard trust in peer-reviewed literature. Considering ongoing technological shifts, this workgroup plans to reconvene in 2 years to assess whether these recommendations should be revised. Given the rapid pace of innovation, the workgroup will actively monitor the AI landscape for emerging technologies and reconvene earlier than 2 years if deemed necessary.

Author affiliations
[1]Anesthesiology, University of Illinois Chicago, Chicago, Illinois, USA
[2]Anesthesia and Pain Management, Toronto Western Hospital and University of Toronto, Toronto, Ontario, Canada
[3]Institute of Health Policy, Management, and Evaluation, University of Toronto, Toronto, Ontario, Canada
[4]Anesthesiology and Pain Medicine, University of Washington, Seattle, Washington, USA
[5]Anesthesiology, University of Pittsburgh Medical Center Health System, Pittsburgh, Pennsylvania, USA
[6]Anaesthesiology, Northern Ontario School of Medicine University, North York, Ontario, Canada
[7]Anesthesiology, University of California San Diego, La Jolla, California, USA
[8]Center for Outcomes Research, UTHealth, Houston, Texas, USA
[9]Anesthesiology, UTHealth, Houston, Texas, USA
[10]Anesthesiology, Vanderbilt University Medical Center, Nashville, Tennessee, USA
[11]Anesthesiology, Perioperative & Pain Medicine, Icahn School of Medicine at Mount Sinai, New York, New York, USA
[12]Anesthesiology, Perioperative and Pain Medicine, Stanford University School of Medicine, Stanford, California, USA
[13]Academic Anaesthesia, Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK
[14]Centre for Applied Health and Social Care Research, Sheffield Hallam University, Sheffield, UK
[15]Anesthesiology, University Hospitals of Leuven, Leuven, Belgium
[16]Cardiovascular Sciences, Biomedical Sciences Group, University of Leuven, Leuven, Belgium
[17]Anesthesiology and Critical Care, University of Pennsylvania, Philadelphia, Pennsylvania, USA
[18]Anesthesiology, Perioperative, and Pain Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA
[19]Anaesthesiology, Intensive Care and Pain Therapy, Saarland University Hospital and Saarland University Faculty of Medicine, Homburg, Germany
[20]Outcomes Research Consortium, Houston, Texas, USA
[21]Anesthesiology, College of Medicine, The Ohio State University, Columbus, Ohio, USA
[22]Anesthesiology, Hospital for Special Surgery, New York, New York, USA
[23]Anesthesiology, Weill Cornell Medicine, New York, New York, USA
[24]Anesthesiology and Perioperative Care Service, VA Palo Alto Health Care System, Palo Alto, California, USA
[25]Anesthesia, Perioperative and Pain Medicine, University of Calgary, South Health Campus, Calgary, Alberta, Canada

Social media Michael R Fettiplace, X @mfettiplace; Anuj Bhatia, X @DrAnujBhatia; Steven L Orebaugh, X @mdorebaugh@twitter.com; Brittani Bungart, X @BrittaniBungart; Christopher P Cheng, X @chrispcheng; Matthew Wiles, X @STHJournalClub; Steve Coppens, X @Steve_Coppens; Vesela P Kovacheva, X @VeselaKovachev2; Christopher L Wu, X @https://twitter.com/ChrisWuMD; Edward R Mariano, X @EMARIANOMD

Acknowledgements Thank you to all the additional members of the AI taskforce who contributed, including Alparslan Turan, Anusha Priya, Chad Brummet, Elena Giovanna Bignami, Franklin Dexter, Graeme McLeod, Guy Weinberg, Jonathan G. Hardman, Kariem El-Boghdadly, Marc Van de Velde, Marco Cascella, Oranicha Jumreornvong, Philip Jones, Rakesh Sondekoppam, Robert Maniker, Romain Pirracchio, Stephan Schwarz, Thomas Hemmerling, and Valentina Bellini.

Collaborators Alparslan Turan, Anusha Priya, Chad Brummet, Elena Giovanna Bignami, Franklin Dexter, Graeme McLeod, Guy Weinberg, Jonathan G. Hardman, Kariem El-Boghdadly, Marc Van de Velde, Marco Cascella, Oranicha Jumreornvong, Philip Jones, Rakesh Sondekoppam, Robert Maniker, Romain Pirracchio, Stephan Schwarz, Thomas Hemmerling, Valentina Bellini.

Contributors MRF, AB, YC, SO, MG, RG, and VI were members of the steering committee who conducted background research and developed the questionnaire. MRF as chair of the workgroup coordinated surveys and feedback and is guarantor of the work. All authors contributed to editing of the manuscript. ChatGPT 4.0 (OpenAI, San Francisco CA, no customizations) was used in the ideation and research for this article (July 2024). ChatGPT o1 (OpenAI, San Francisco CA, February and March 2025) was used for help in drafting of the introduction and discussion. Grammarly was used for help in grammar and fluency (Grammarly, San Francisco, CA. February 2025). Anuj Bhatia, Yian Chen, Steven Orebaugh, Michael Gofeld, Rodney A Gabriel, and Vivian Ip are Steering committee members.

Funding Dr. Joseph was supported by NIH grant K08GM139031 from NIGMS, Dr. Lonsdale is supported by NIH grant 5T32GM108554. Dr. Kovacheva reports funding from NIH/NHLBI grant 1K08HL161326."

Disclaimer The author is a staff member of the World Health Organization. The author alone is responsible for the views expressed in this publication and they do not necessarily represent the views, decisions or policies of the World Health Organization.

Competing interests None declared.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer-reviewed.

Data availability statement Data are available upon reasonable request. Summary response data is available upon reasonable request.

ORCID iDs
Michael R Fettiplace http://orcid.org/0000-0001-6946-8355
Anuj Bhatia http://orcid.org/0000-0001-6292-8654
Yian Chen http://orcid.org/0000-0002-5971-9111
Steven L Orebaugh http://orcid.org/0000-0001-9517-6575
Michael Gofeld http://orcid.org/0000-0001-7372-4171
Rodney A Gabriel http://orcid.org/0000-0003-4443-0021
Daniel I Sessler http://orcid.org/0000-0001-9932-3077
Hannah Lonsdale http://orcid.org/0000-0001-9953-8844
Brittani Bungart http://orcid.org/0000-0001-5705-3742
Christopher P Cheng http://orcid.org/0000-0001-9112-9464
Garrett W Burnett http://orcid.org/0000-0003-0565-9499
Matthew Wiles http://orcid.org/0000-0003-2299-0060
Steve Coppens http://orcid.org/0000-0002-6026-7757
Kristin L Schreiber http://orcid.org/0000-0002-7361-2214
Thomas Volk http://orcid.org/0000-0001-5175-7159
Christopher L Wu http://orcid.org/0000-0002-4484-0787
Edward R Mariano http://orcid.org/0000-0003-2735-248X
Vivian H Y Ip http://orcid.org/0000-0001-6158-4415

REFERENCES

1 OpenAI. ChatGPT 4.0. Available: https://chatgpt.com/ [Accessed 2 Jun 2024].
2 Anthorp\c. claude. Available: https://claude.ai/login?returnTo=%2F%3F [Accessed 2 Jun 2024].
3 Google. Google gemini. Available: https://gemini.google.com/ [Accessed 2 Jun 2024].
4 Thorp HH. ChatGPT is fun, but not an author. *Science* 1979;379:313.
5 Hadi MU, Tashi QA, Shah A, *et al*. Large language models: a comprehensive survey of its applications, challenges, limitations, and future prospects. Techrxiv (pre-print) [Preprint] 2024.
6 Lonsdale H, O'Reilly-Shah VN, Padiyath A, *et al*. Supercharge Your Academic Productivity with Generative Artificial Intelligence. *J Med Syst* 2024;48:73:1–7:.
7 Májovský M, Černý M, Kasal M, *et al*. Artificial Intelligence Can Generate Fraudulent but Authentic-Looking Scientific Medical Articles: Pandora's Box Has Been Opened. *J Med Internet Res* 2023;25:e46924.
8 Anderson N, Belavy DL, Perle SM, *et al*. AI did not write this manuscript, or did it? Can we trick the AI text detector into generated texts? The potential future of ChatGPT and AI in Sports & Exercise Medicine manuscript generation. *BMJ Open Sport Exerc Med* 2023;9:e001568.

9 Hernandez-Boussard T, Bozkurt S, Ioannidis JPA, *et al*. MINIMAR (MINimum Information for Medical AI Reporting): Developing reporting standards for artificial intelligence in health care. *J Am Med Inform Assoc* 2020;27:2011–5.
10 Vasey B, Nagendran M, Campbell B, *et al*. Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *BMJ* 2022;377:e070904.
11 Liu X, Cruz Rivera S, Moher D, *et al*. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nat Med* 2020;26:1364–74.
12 Wu CL, Cho B, Gabriel R, *et al*. Addition of dexamethasone to prolong peripheral nerve blocks: a ChatGPT-created narrative review. *Reg Anesth Pain Med* 2024;49:777–81.
13 Wallace E, Feng S, Kandpal N, *et al*. Universal adversarial triggers for attacking and analyzing nlp. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP); Stroudsburg, PA, USA, Hong Kong, China. 10.18653/v1/D19-1221 Available: https://www.aclweb.org/anthology/D19-1
14 Sessler DI, Turan A. Journal policy on large language generative models. *J Clin Anesth* 2024;96:111450.
15 Flanagin A, Pirracchio R, Khera R, *et al*. Reporting Use of AI in Research and Scholarly Publication-JAMA Network Guidance. *JAMA* 2024;331:1096–8.
16 Nasa P, Jain R, Juneja D. Delphi methodology in healthcare research: How to decide its appropriateness. *World J Methodol* 2021;11:116–29.
17 Bajwa J, Munir U, Nori A, *et al*. Artificial intelligence in healthcare: transforming the practice of medicine. *Future Healthc J* 2021;8:e188–94.
18 Kathryn F. *The Rand/UCLA appropriateness method user's manual*. Rand 2001.
19 Li H, Moon JT, Purkayastha S, *et al*. Ethics of large language models in medicine and medical research. *Lancet Digit Health* 2023;5:e333–5.
20 Goldberg CB, Adams L, Blumenthal D, *et al*. To do no harm - and the most good - with AI in health care. *Nat Med* 2024;30:623–7.
21 Open evidence. Available: https://www.openevidence.com/ [Accessed 11 Jun 2025].
22 Bowness JS, Burckett-St Laurent D, Hernandez N, *et al*. Assistive artificial intelligence for ultrasound image interpretation in regional anaesthesia: an external validation study. *Br J Anaesth* 2023;130:217–25.
23 Chen Z, Chen C, Yang G, *et al*. Research integrity in the era of artificial intelligence: Challenges and responses. *Medicine (Abingdon)* 2024;103:e38811.
24 Jairoun AA, El-Dahiyat F, ElRefae GA, *et al*. Detecting manuscripts written by generative AI and AI-assisted technologies in the field of pharmacy practice. *J Pharm Policy Pract* 2024;17:2303759.
25 Abd-Alrazaq A, AlSaad R, Alhuwail D, *et al*. Large Language Models in Medical Education: Opportunities, Challenges, and Future Directions. *JMIR Med Educ* 2023;9:e48291.
26 Biswas S. ChatGPT and the Future of Medical Writing. *Radiology* 2023;307:e223312:1–3:.
27 Editorial N. Tools such as ChatGPT threaten transparent science; here are our ground rules for their use. *Nature New Biol* 2023;613:612.
28 The Lancet Digital Health. *Lancet Digit Health* 2023;5:e102.
29 Elsevier. Generative ai policies for journals. 2024. Available: https://www.elsevier.com/about/policies-and-standards/generative-ai-policies-for-journals#1-for-authors [Accessed 26 Jan 2025].
30 International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals 2024.
31 Boiko DA, MacKnight R, Kline B, *et al*. Autonomous chemical research with large language models. *Nature New Biol* 2023;624:570–8.
32 Lu C, Lu C, Lange RT, *et al*. The AI Scientist: Towards Fully Automated Open-Ended Scientific Discovery. *ArxivOrg (Pre-Print)* 2024;12.
33 The Lancet. Information for Authors. *The Lancet* 2025.
34 Reproducibility and Replicability in Science. Washington, D.C: National Academies Press, 2019.
35 Ong JCL, Chang S-H, William W, *et al*. Medical Ethics of Large Language Models in Medicine. *NEJM AI* 2024;1:1–11.
36 Anom BY. Ethics of Big Data and artificial intelligence in medicine. *Ethics Med Public Health* 2020;15:100568.
37 Mello MM, Guha N. Understanding Liability Risk from Using Health Care Artificial Intelligence Tools. *N Engl J Med* 2024;390:271–8.
38 Pressman SM, Borna S, Gomez-Cabello CA, *et al*. AI and Ethics: A Systematic Review of the Ethical Considerations of Large Language Model Use in Surgery Research. *Healthcare (Basel)* 2024;12:825.
39 Chetwynd E. Ethical Use of Artificial Intelligence for Scientific Writing: Current Trends. *J Hum Lact* 2024;40:211–5.
40 o3 Deep Research. Revisiting the mckinley tariff of 1890 through the lens of modern trade theory. 2025.
41 Grammerly. Available: https://www.grammarly.com/ [Accessed 3 Jun 2024].
42 Murdoch B. Privacy and artificial intelligence: challenges for protecting health information in a new era. *BMC Med Ethics* 2021;22:122.
43 National institutes of health. nih scientific data sharing. Available: https://sharing.nih.gov/data-management-and-sharing-policy/about-data-management-and-sharing-policies/data-management-and-sharing-policy-overview#after [Accessed 2 Jun 2024].
44 NIH supported data sharing resources. 2025. Available: https://www.nlm.nih.gov/NIHbmic/domain_specific_repositories.html [Accessed 14 Feb 2025].
45 Hrynaszkiewicz I, Altman DG. Towards agreement on best practice for publishing raw clinical trial data. *Trials* 2009;10:17:1–5:.
46 Lapid MI, Clarke BL, Wright RS. Institutional Review Boards: What Clinician Researchers Need to Know. *Mayo Clin Proc* 2019;94:515–25.
47 COPE Coucil. Authorship and AI tools. 2024.
48 World Association of Medical Editors. Chatbots, Generative AI, and Scholarly Manuscripts: WAME Recommendations on Chatbots and Generative Artificial Intelligence in Relation to Scholarly Publications. *Philippine Journal of Otolaryngology-Head and Neck Surgery* 2023;38:7–9.
49 Flanagin A, Bibbins-Domingo K, Berkwits M, *et al*. Nonhuman "Authors" and Implications for the Integrity of Scientific Publication and Medical Knowledge. *JAMA* 2023;329:637.
50 Chemaya N, Martin D. Perceptions and detection of AI use in manuscript preparation for academic journals. *PLoS ONE* 2024;19:e0304807.
51 DHHS. Office of the secretary ethical principles and guidelines for the protection of human subjects of research the national commission for the protection of human subjects of biomedical and behavioral research action: notice of report for public comment. 1979.