Impact Surgery

Original Research Paper

e-ISSN: 3033-4470

Generative Artificial Intelligence Transparency in scientific writing: the GAIT 2024 guidance

GAIT 2024 Collaborative Group*

Correspondence: Dr. Dmitri Nepogodiev, Department of Applied Health Sciences, Institute of Translational Medicine, Mindelsohn Way, Birmingham B15 2TH, UK, d.nepogodiev@bham.ac.uk

Abstract

Background: Generative Artificial Intelligence (GAI) tools are increasingly used in research. At present, there is no standardised approach to reporting GAI use. We aimed to produce guidance to support authors in the use of GAI in scientific writing.

Methods: A steering group of academic surgeons with experience in GAI developed draft statements for best practice in reporting GAI use. These statements were refined through iterative discussions using a nominal group technique. A broad network of surgeons and surgical researchers were invited to participate in an online consultation exercise to validate these statements by ranking using a Likert scale. A pre-planned threshold of ≥70% of participants scoring a statement ≥7 would lead to acceptance. Participants were additionally surveyed on the use, opportunities, and risks. Thematic analysis was completed using ChatGPT.

Results: The steering group developed five draft statements, which were validated in the online consultation exercise by 124 participants from 46 countries. Four draft statements were accepted based on this exercise and consolidated into the final Generative AI Transparency (GAIT) guidance: (1) GAI use should be reported in a GAIT statement; (2) GAI use should be mapped using the Contributor Roles Taxonomy; (3) specific prompts used should be reported; (4) authors should retain final responsibility for their work. Example statements to be included in manuscripts include: (1) ChatGPT-4o was used in November 2024 to check and edit statistical code (formal analysis) and edit small sections of the manuscript text for clarity (writing: review & editing). Prompts used are reported in the supplement. The authors should retain final responsibility for their work; (2) No Generative Artificial Intelligence was used to produce, draft, or edit this guidance paper.

Conclusion: The GAIT 2024 guidance will support transparent, structured reporting of the use of generative AI in scientific writing, supporting the integrity of research outputs.

**Members of the GAIT 2024 Collaborative Group are co-authors of this study and are listed under Appendix E.*

Cite as: GAIT 2024 Collaborative Group. Generative Artificial Intelligence Transparency in scientific writing: the GAIT 2024 guidance. *Impact Surgery*, 2(1), 6-11. https://doi.org/10.62463/surgery.134

Introduction

Generative Artificial Intelligence (GAI) tools are rapidly increasing in scale and adoption[1]. These innovative tools are currently being explored by researchers to aid in various stages of the scientific writing process[2]. Most widely used are Large Language Models (LLMs), which uses machine learning techniques such as recurrent neural networks to train on vast amounts of data[3]. These models are able to comprehend text inputs and produce human-like responses[4]. LLM use has been reported throughout the research process, including conceptualisation, qualitative data synthesis, summarisation, and academic writing[5]. Other models, such as Generative Adversarial Networks and AutoML Platforms, can generate medical images and help with medical data synthesis and analysis[6]. Using GAI tools may increase the speed of research as it can automate repetitive tasks and enable more complex and sophisticated data analysis. Systematic review GAI software, such as Elicit, uses LLM to automate literature searches, summaries and data extraction, and could transform the systematic review process[7].

Some scientific journals have guidelines on the use and reporting of GAI in medical research, but there is a considerable variation[8, 9]. The International Journal Committee of Medical Editors recommend recognising and acknowledging the contribution of artificial intelligence, including full citations, but stop short of requiring specific statements[10]. The Equator Network has several guidelines on the use and reporting of artificial intelligence in research[11]. However, there are no universally accepted standards guiding the formal use and accreditation of GAI in medical research[12]. Surgical research teams urgently require practical guidance in transparent reporting of GAI use in academic practice. This study aimed to present consensus-based guidance for how to report use of GAI in scientific writing.

Methods:

Draft statements of best practice

A diverse steering group was formed, that included academic surgeons, public health specialists, statisticians, and methodologists. Each steering group member had extensive prior experience of using GAI, scientific writing, and publishing. Of the 7 steering group members, 43% were female and 43% identified as belonging to minority ethnic groups. The steering group developed statements of best practice relating to the use of GAI in scientific writing using nominal group technique[13]. Members individually drafted statements which were then discussed by the group, iteratively refined and ranked. A set of draft statements were synthesised that represented the steering group's consensus.

Online consultation exercise

The draft statements were validated through an online consultation exercise with surgeons and surgical researchers. Participants were identified using a snowball sampling method, starting from an online community of *Impact Surgery* readers and reviewers. Participants were invited to complete a short online survey on 19-23 April 2024. In the survey they rated their agreement with each draft statement on a Likert scale from 1 (strongly disagree) to 10 (strongly agree).

It was pre-planned that a statement would be considered to have been accepted if at least 70% of participants scored it ≥7 on the Likert scale. In addition, average scores were calculated for each statement. The survey was conducted using the Research Electronic Data Capture web application (REDCap, Vanderbilt University, Nashville, TN, USA), hosted by the University of Birmingham, UK.

Thematic analysis

To explore how GAI can be used in scientific writing, survey participants were asked two free-text questions "If you have, please describe how you used generative AI to help you to produce one or more surgical research outputs" and "What do you think are the opportunities that generative AI offers for surgical research?". To identify potential ethical considerations around the use of generative AI in surgical research, the following free-text question was included "What do you think are the risks of using generative AI offers for surgical research?". Responses to these questions were analysed by prompting ChatGPT-4 to identify emergent themes (Appendix B-D). The outputs were reviewed by two authors to ensure meaningful content and to screen any hallucinations.

Final guidance

The steering group considered the results of the online consultation exercise, including additional insights from the thematic analysis. A nominal group technique was again used to consolidate the draft statements that were successfully validated in the online consultation exercise in to streamlined, final guidance.

Ethics

Completion of the UK Health Research Authority's ethics self-assessment tool (https://www.hra-decisiontools.org.uk/research/) confirmed that this study did not require formal ethical approval.

Results:

Draft statements for best practice

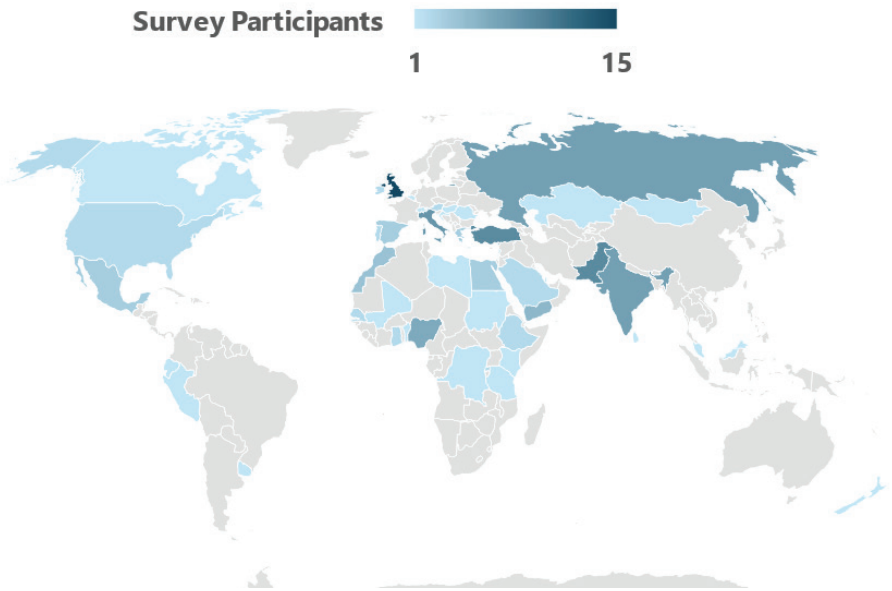
The expert steering group developed five key recommendations (Table 1). The Contributor Role Taxonomy (CRediT) was proposed as a framework to map the use of GAI. CRediT is an established set of 14 authorship roles that is used to characterise authors' contributions to research outputs[14].

Online consultation exercise

A total of 124 surgeons and surgical researchers participated in the online consensus exercise, representing 46 countries (Figure 1). Twenty-four (19.4%) of participants had not published any surgical research papers, 43 (35.7%) had published up to nine papers, 34 (27.4%) had published 10-49 papers, and 23 (18.5%) had published 50 or more papers. Over a third (35.5%, 44/124) had used GAI to support the production of one or more surgical research outputs. A further 50% (62/124) had not previously used GAI for surgical research, but planned to use it in future studies.

Four draft statements were accepted in the online consultation exercise; but the statement relating to discussion of the strengths and weaknesses of GAI use failed to attract sufficient support. All statements achieved average scores ≥7 (Table 1).

Figure 1. Geographical Distribution of Survey Responses



Thematic Analysis

The key ethical considerations that the online consultation exercise identified were data privacy, data interpretation and quality assurance (Table 2). Respondents with previous experience of GAI reported uses for writing assistance, literature review and statistical analysis (Appendix B). Opportunities for GAI in surgical research included improved efficiency, training, personalised medicine and robotic surgery (Appendix C).

Final recommendations

The steering group synthesised the online consultation

Table 1: Online consultation exercise results

Draft statement for best practice	Average score	Participants score statement ≥7
1. If generative AI has been used in a research output, this should be reported (including the model (e.g. ChatGPT, Gemini, DALL-E) and version used).	8	81
2. If generative AI has been used in a research output, this should be declared in a "generative AI usage statement". If generative AI has not been used this should be stated.	8	73
3. If generative AI has been used in a research output, the CRediT taxonomy should be used to map the use of generative AI.	8	78
4. If generative AI has been used in a research output, the prompts used should be listed in a supplementary file.	8	70
5. If generative AI has been used in a research output, the strengths and weaknesses of this approach should be reported.	8	69

Table 2: Ethical considerations in the use of generative AI in surgical research

Data privacy
• How is data entered as prompts into the generative AI stored? • Is the confidentiality of data entered as prompts into the generative AI assured? • Does the generative AI 'learn' from the data entered as prompts?
Data interpretation
• How can researchers minimise the risk of bias because of generative AI reproducing biases in training datasets? • Is there a risk of reduced originality and innovation if researchers use generative AI to develop research ideas and methodologies?
Quality assurance
• How can researchers ensure reproducibility of methods (e.g. the same prompt may produce different results when a query is repeated)? • How can researchers verify the accuracy of information produced by generative AI (i.e. to identify and address 'hallucinations')? • How can researchers validate proposed methodologies if they are not experienced in these? • If there is insufficient quality assurance of generative AI supported research, could this further increase the production of low-quality or fraudulent research, with the potential to harm patients? • Is the current peer-review process equipped to adequately evaluate GAI-assisted research? • How can the peer-review process be improved for GAI-assisted research, since AI detectors are limited to their access to specific LLMs and hence are likely to always lag behind? • Could there be a risk of skill degradation as researchers become increasingly reliant on generative AI? • Could GAI have a role in overcoming language barriers with allowing the scientific community using English language to interact with their non-English using counterparts and if so, how would peer review process be impacted?

exercise findings and produced the final Generative AI Transparency (GAIT) guidance:

- GAI use should be reported in a GAIT statement, including the model and version used, and the date the GAI was last used.
- GAI use should be mapped using the CRediT criteria. To demonstrate how the CRediT criteria could be used in this context, potential GAI use cases have been mapped against CRediT in Table 3.
- Specific prompts used should be reported in a supplement.
- Authors should retain final responsibility for their work.

GAIT statements could be placed in either a dedicated acknowledgment section or the methods. Example GAIT statements are in Table 4.

Discussion:

This paper provides initial guidance on how to acknowledge use of GAI in scientific writing. This study is unique in consolidating opinion on GAI from clinicians and academics with a wide range of experience across diverse geographic setting.

There is broad consensus within the academic community about the need to report the use of GAI within the paper[12, 15-17], but many leading publishers do not yet have GAI guidelines[12]. Among those that do, there is substantial heterogeneity in reporting recommendations; many guidelines do not require reporting of prompts[16], there is no consensus about where in the manuscript GAI contributions should be reported[12], and there are no existing frameworks for reporting GAI contributions. This existing variability in GAI guidelines diminishes efficacy, creates confusion and places additional challenges for researchers[12]. Given that the field of GAI is rapidly

Table 3: Generative artificial intelligence use cases in scientific writing mapped against CRediT criteria

Contributor role	Examples
Conceptualization	• Interactively discussing and refining potential research ideas. • Producing a summary of existing knowledge from online resources.
Data curation	• Automating aspects of tasks such as data cleaning (identifying data outliers, recoding data variables).
Formal Analysis	• Supporting the development of an appropriate statistical analysis plan. • Generating code to support statistical analysis. • Supporting advanced machine learning analyses.
Funding acquisition	• Generating an initial draft and/or refining a research grant application.
Investigation	• Extracting study data from larger, machine-readable datasets. • Performing basic qualitative analysis of interview transcript data.
Methodology	• Interactively discussing and refining methodology, identifying and addressing potential methodological weaknesses.
Project administration	• Generating copy for use in project documentation, marketing, and newsletters, etc.
Software	• Generating code to produce data capture apps.
Visualisation	• Suggesting possible approaches to graphical presentation of complex data. • Generate code to support data visualisation. • Generating visualisations.
Writing – original draft	• Helping authors to plan the outline for their manuscript. • Generating an initial draft based on prompt outlining key concepts for introduction and discussion.
Writing – review & editing	• Revising sections of authors' draft for succinctness, clarity, or professionalism.

**Large language models are less likely to directly feed into the Resources, Supervision, Validation contributor roles.*

progressing, there is an urgent need to standardise guidelines and provide authors with a framework for reporting.

The GAIT guidance is consistent with existing journal and publishing group guidelines, but substantially improves the depth and consistency of reporting of GAI use; for example, the recommendation to map GAI use against CRediT criteria is novel. Using our consensus-driven guidance, publishers, journals, and research organisations can structure more harmonized guidelines.

In addition to providing a consensus statement, our consultation exercise highlighted the risks and benefits of GAI in scientific writing. It raised concerns about the accuracy of GAI output; for example, it is not currently possible to tell if ChatGPT is "hallucinating" other than cross checking results[18]. However, GAIs are likely to improve over time, reducing inaccuracies. There were additional concerns raised about the privacy of data uploaded to GAI models, which is an ongoing debate both within and outside of academia[1]. Another concern identified was the inherent bias within the model, learned during training from online content[18]. There are further considerations about how peer review can match the pace of GAI development. There is no validated way of identifying content plagiarised by unreported GAI use19. Many participants highlighted the benefits of GAI in research, envisioning a role for GAI to contribute to

Table 4: Example Generative Artificial Intelligence Transparency (GAIT) statements

Example statement	Context
ChatGPT-4o was used in November 2024 to check and edit statistical code (formal analysis) and edit small sections of the manuscript text for clarity (writing – review & editing). Prompts used are reported in the supplement. The authors should retain final responsibility for their work.	Reports use of GAI
No Generative Artificial Intelligence was used to produce, draft, or edit this guidance paper.	Reports GAI not used

manuscript writing, data analysis, systematic review. Using GAI to easily create videos and three-dimensional animations can help increase the impact of research and improve research community engagement. Overall this could increase the speed, efficiency and equality of research, increasing uptake of GAI use in academic medicine[20].

This study has limitations. This survey was developed by surgical authors for surgical research, so elements may not be appropriate for other disciplines. Most respondents were from Europe, so further work is needed to involve researchers using GAI in LMICs. Finally, almost 20% of respondents had not previously published academic papers, so their responses were grounded in less research experience.

It is highly likely further reporting guidelines will be produced in the future, including those from the CANGARU group[8]. GAIT 2024 provides immediate guidance and it is likely that it will need to flexibly evolve as GAI use in scientific writing develops further, for example, becoming more common in complex tasks such as data manipulation and AI integration. Further work to evaluate and revise the GAI statement would ensure it remains relevant and contemporaneous. Additionally, the role of GAI in the peer review process warrants exploration.

Conflicts of interest: None declared

References

1. Morley J, DeVito NJ, Zhang J. Generative AI for medical research. *Br Med J.* 2023;382:p1551.

2. Zhang P, Kamel Boulos MN. Generative AI in Medicine and Healthcare: Promises, Opportunities and Challenges. *Future Internet.* 2023;15. doi: 10.3390/fi15090286.

3. Aydın Ö, Karaarslan E. OpenAI ChatGPT Generated Literature Review: Digital Twin in Healthcare. *SSRN Electron J.*2022.

4. Wu T, et al. A Brief Overview of ChatGPT: The History, Status Quo and Potential Future Development. *IEEE/CAA J Autom Sin.* 2023;10(5):1122-1136.

5. Omiye JA, et al. Large Language Models in Medicine: The Potentials and Pitfalls. *Ann Intern Med.* 2024;177(2):210-220.

6. Shin H-C, et al. Medical Image Synthesis for Data Augmentation and Anonymization Using Generative Adversarial Networks. In: *Springer International Publishing*; 2018. p. 1-11.

7. Kung J. Elicit (product review). *J Can Health Libr Assoc.* 2023;44(1).

8. Cacciamani G, et al. Development of the ChatGPT, Generative Artificial Intelligence and Natural Large Language Models for Accountable Reporting and Use (CANGARU) Guidelines. 2023.

9. COPE Council. COPE position - Authorship and AI. Accessed Jan 2025. Available from: https://publicationethics.org/guidance/cope-position/authorship-and-ai-tools

10. International Committee of Medical Journal Editors (ICMJE). Defining the Role of Authors and Contributors. 2024 [cited 2024 Oct 17]. Available from: https://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html.

11. Olczak J, et al. Presenting artificial intelligence, deep learning, and machine learning studies to clinicians and healthcare stakeholders: an introductory reference with a guideline and a Clinical AI Research (CAIR) checklist proposal. *Acta Orthop.* 2021;92(5):513-525.

12. Ganjavi C, et al. Publishers' and journals' instructions to authors on use of generative artificial intelligence in academic and scientific publishing: bibliometric analysis. *Br Med J.* 2024:e077192.

13. Van de Ven AH, Delbecq AL. The nominal group as a research instrument for exploratory health studies. *Am J Public Health.* 1972;62(3):337-342.

14. Alpi KM, Akers KG. CRediT for authors of articles published in the Journal of the Medical Library Association. *J Med Libr Assoc.* 2021;109(3):362-364.

15. Hamm B, Marti-Bonmati L, Sardanelli F. ESR Journals editors' joint statement on Guidelines for the Use of Large Language Models by Authors, Reviewers, and Editors. *Eur Radiol Exp.* 2024;8(1).

16. Kaebnick GE, et al. Editors' statement on the responsible use of generative AI technologies in scholarly journal publishing. *Med Health Care Philos.* 2023;26(4):499-503.

17. Leung TI, et al. Best Practices for Using AI Tools as an Author, Peer Reviewer, or Editor. *J Med Internet Res.*2023;25:e51584.

18. Bhattacharyya M, et al. High Rates of Fabricated and Inaccurate References in ChatGPT-Generated Medical Content. *Cureus.* 2023;15(5):e39238.

19. Habibzadeh F. Plagiarism: A Bird's Eye View. *J Korean Med Sci.* 2023;38(45):e373.

20. Liebrenz M, et al. Generating scholarly content with ChatGPT: ethical challenges for medical publishing. *Lancet Digit Health.* 2023;5(3):e105-e106.