

Academic Writing with Artificial Intelligence in Neurosurgery

J.Sales-Llopis

Neurosurgery Department, *General University Hospital Alicante, Spain*

Abstract

Artificial Intelligence (AI) is rapidly transforming academic writing, particularly in specialized fields such as neurosurgery. Large language models (LLMs) like ChatGPT offer practical benefits including improved readability, support for non-native English speakers, assistance in drafting abstracts and introductions, and tools for visualizing data and organizing results. Recent studies (Fauziah et al., Schneider et al., Akgun et al., Nabata et al.) demonstrate that AI-generated content often matches or exceeds human-authored texts in surface-level quality and readability, though it may lack depth, originality, and clinical nuance. Despite their potential, AI tools raise ethical concerns regarding authorship attribution, misinformation, and data privacy, especially in contexts involving patient information. Studies reveal that human reviewers frequently fail to distinguish between AI- and human-generated writing and sometimes prefer AI outputs for their clarity. The growing presence of AI in neurosurgical publishing underscores the need for transparency, ethical guidelines, and integration policies. While AI cannot replace expert insight, it serves as a powerful augmentation tool that—when used responsibly—enhances efficiency, clarity, and accessibility in scientific communication.

Keywords: Artificial Intelligence · Large Language Models · Neurosurgery · Academic Writing · Scientific Publishing · Authorship Ethics · Readability · ChatGPT

Introduction

Academic writing plays a central role in the dissemination of knowledge, advancement of clinical practice, and development of evidence-based medicine in neurosurgery. However, the increasing demands of research productivity, publication pressure, and the need for high-quality English prose—especially among non-native speakers—pose significant challenges for researchers. In this evolving landscape, Artificial Intelligence (AI), particularly large language models (LLMs) such as ChatGPT, is emerging as a powerful tool to assist in the generation, refinement, and structuring of academic content.

AI-based writing assistants offer a range of capabilities, from correcting grammar and enhancing clarity to generating full drafts of scientific abstracts, introductions, and technical reports. Their integration into the academic workflow can streamline manuscript preparation, support literature review processes, and improve accessibility and readability of complex scientific information. This is especially relevant in neurosurgery, where the high technical demands of the discipline often

intersect with the need for precise and nuanced communication.

Despite these advantages, the use of AI in academic writing raises critical questions around originality, accuracy, authorship attribution, and ethical use. Studies have shown that AI-generated texts can be indistinguishable from human-authored ones and are often preferred for their readability, yet concerns remain regarding their depth of analysis, clinical judgment, and factual reliability.

This article explores the opportunities and limitations of AI in neurosurgical academic writing, drawing upon recent comparative studies, including works by Fauziah et al., Schneider et al., Akgun et al., and Nabata et al. Through a synthesis of current evidence and ethical considerations, we aim to provide a balanced perspective on how AI tools can be responsibly integrated into the scientific publishing process—complementing rather than replacing human expertise.

Materials and Methods

1973-1974

This article adopts a narrative review approach to examine the role of Artificial Intelligence (AI), particularly large language models (LLMs), in academic writing within the field of neurosurgery. The objective is to synthesize recent literature, highlight emerging trends, and critically assess both the potential and limitations of AI-assisted writing tools in scientific publishing.

To inform this analysis, we conducted a targeted literature search in PubMed, Scopus, and Google Scholar for studies published between January 2023 and March 2025. Search terms included combinations of “artificial intelligence”, “large language models”, “ChatGPT”, “academic writing”, “scientific publishing”, “neurosurgery”, and “authorship ethics”. Priority was given to peer-reviewed studies that directly compared AI-generated and human-authored texts, assessed readability metrics, or explored the ethical and practical implications of AI use in manuscript preparation.

Four key studies were selected for in-depth analysis:

Fauziah et al. (2025) – Comparative readability and perception of AI vs. human-written neurosurgery articles
Schneider et al. (2025) – Detection and prevalence of AI-generated content in neurosurgical journals
Akgun et al. (2025) – Expert evaluation of AI-generated vs. human-authored full-length neurosurgery manuscripts
Nabata et al. (2025) – Accuracy of human reviewers in identifying AI-generated scientific abstracts
In addition to these core studies, supplemental literature was reviewed to contextualize findings within broader discussions on AI ethics, transparency in authorship, and academic integrity. Where relevant, observations from clinical and educational use cases of AI-assisted writing tools were integrated to provide practical insights for neurosurgeons, trainees, and academic institutions.

No patient data were used in this review. The analysis is based solely on publicly available literature and does not require ethics committee approval.

Results

Four primary studies were identified that directly investigated the role and impact of AI-generated content in neurosurgical academic writing. These studies collectively highlight a shift toward the

integration of large language models (LLMs) such as ChatGPT in various stages of manuscript development, from abstract drafting to full-length article generation.

Readability and Preference: Both Fauziah et al. and Akgun et al. reported higher readability scores for AI-generated texts compared to those written by humans. In both studies, participants showed similar or even higher preference for AI-generated content (47–63%), despite modest accuracy in detecting AI authorship (around 61%).

Prevalence and Detection: Schneider et al. found that approximately 20% of recent neurosurgical publications contained sections likely generated by AI, especially in the methods and abstract sections. However, after removing abstracts from the analysis, the statistical significance of section-specific AI prevalence was lost, suggesting inherent formatting biases in detection tools.

Reviewer Discrimination: Nabata et al. demonstrated that most surgical faculty and trainees could not reliably distinguish between original and AI-generated abstracts. Surprisingly, a significant portion (63.4%) preferred the AI-written abstract, emphasizing the persuasive clarity and structure of LLM outputs.

Ethical Implications: All four studies underline the urgent need for transparency in AI use. While the benefits of AI are clear—particularly in enhancing clarity and reducing language barriers—issues surrounding factual accuracy, authorship credit, and ethical oversight remain unresolved.

Study	Study Type	Main Findings	Implications
Fauziah et al. (2025)	Comparative analysis of AI vs. human-written articles	AI-generated texts had higher readability but less depth; 61% accuracy in identifying AI authorship	AI improves accessibility but lacks clinical nuance; requires expert oversight
Schneider et al. (2025)	Prevalence study using AI-detection tools	20% of neurosurgical manuscripts contained AI-generated content; especially in abstracts and methods	AI detection is increasingly unreliable; ethical disclosure and transparency are essential
Akgun et al. (2025)	Expert evaluation of article quality	ChatGPT-generated articles had higher readability and comparable quality; 47% preferred AI, 44% human	AI can assist in drafting; fact-checking and clinical validation are critical
Nabata et al. (2025)	Observational study with surgical faculty and trainees	Only 40% correctly identified human-written abstract; 63.4% preferred the AI-generated version	Human reviewers struggle to distinguish AI texts; underscores need for transparency and guidelines

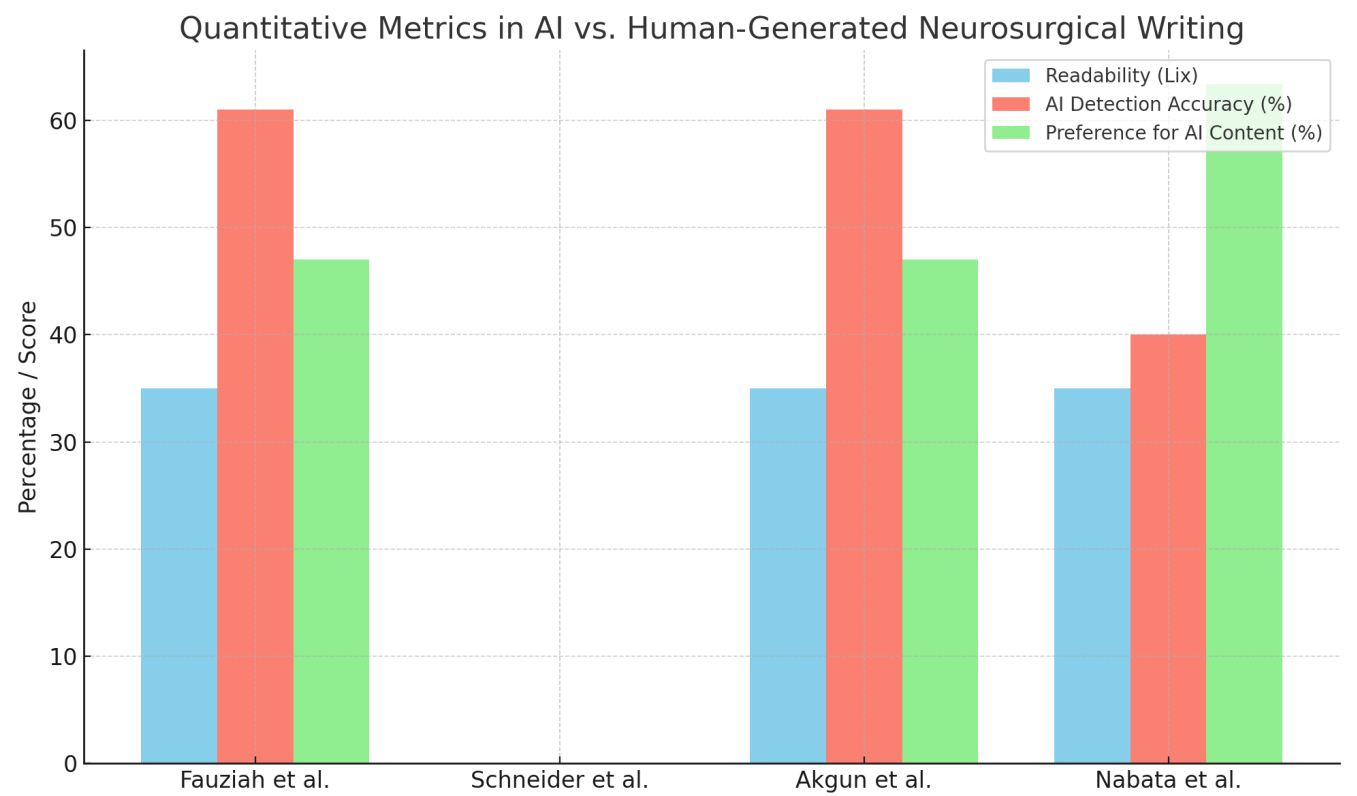


Figure 1. Quantitative Metrics in AI vs. Human-Generated Neurosurgical Writing. This bar chart summarizes findings from four recent studies analyzing AI-generated content in neurosurgical academic writing. Metrics include Lix readability scores, percentage of accurate identification of AI-generated texts, and user preference for AI-generated content. The results highlight the consistently higher readability and significant preference for AI-written content, while also illustrating challenges in reliably detecting AI authorship. Data from Schneider et al. are partial due to lack of reported values for certain metrics.

Discussion

The findings from this review illustrate both the transformative potential and inherent limitations of Artificial Intelligence (AI), particularly large language models (LLMs) like ChatGPT, in academic neurosurgical writing. Across the analyzed studies, AI-generated content demonstrated significantly higher readability and, in many cases, equal or greater user preference when compared to traditional human-authored texts. These results suggest that AI tools can effectively streamline the writing process, especially for tasks requiring formal structure and linguistic clarity, such as abstracts, methods sections, and technical notes.

One of the most striking observations is the difficulty experts face in distinguishing between AI- and human-authored content. In studies by Fauziah et al. and Nabata et al., reviewers correctly identified AI-generated content only 40–61% of the time—barely above chance. Furthermore, the AI-generated abstracts were preferred by a substantial portion of respondents, underscoring the persuasive power of AI in producing clear and engaging prose. These findings suggest that AI may contribute to democratizing access to academic publishing, particularly for non-native English speakers and early-career researchers who may struggle with linguistic and stylistic barriers.

However, this growing integration of AI into academic writing raises complex ethical and methodological concerns. AI tools can “hallucinate” references, fabricate plausible-sounding facts,

and omit clinical nuance—shortcomings that are particularly problematic in a discipline like neurosurgery, where precision and context are paramount. The use of AI in drafting manuscripts without sufficient human oversight could compromise the quality and credibility of scientific communication.

Moreover, the studies highlight limitations in current detection methods. As shown by Schneider et al., even advanced AI-detection strategies using tools like RoBERTa combined with perplexity thresholds struggle to maintain consistency and reliability. This suggests that efforts to “police” AI usage may be less effective than developing robust policies for ethical disclosure and proper attribution of AI-generated contributions.

The current binary classification of authorship—AI-generated or human-authored—also fails to account for the reality of hybrid writing processes. Most researchers using AI do so as co-authors or assistants rather than full content generators. Thus, future research and editorial guidelines should explore more nuanced frameworks that recognize degrees of AI involvement.

Another overlooked but important benefit is the potential educational value of AI tools. When used transparently and responsibly, LLMs can serve as writing tutors for students and residents, simulate peer review processes, and assist in literature synthesis and concept clarification. Their integration into medical education could foster greater engagement with academic writing and reduce disparities in publication opportunities.

Finally, the use of readability metrics, while helpful in measuring surface-level clarity, should not be conflated with scientific rigor or depth. As some authors have cautioned, overly simplified text may compromise the precision and complexity needed for effective scholarly communication. Readability should therefore be viewed as a complement—not a substitute—for expert-driven content creation.

Conclusion

Artificial Intelligence, and specifically large language models like ChatGPT, is poised to reshape academic writing in neurosurgery. By enhancing readability, supporting non-native English speakers, and streamlining the drafting of scientific texts, AI offers practical benefits that can improve accessibility and efficiency in research communication. However, its limitations—particularly in clinical accuracy, depth of analysis, and ethical attribution—underscore the need for cautious and transparent integration.

The current evidence suggests that AI should not be viewed as a replacement for expert authorship, but rather as a complementary tool that augments human capability. Ongoing efforts must focus on establishing ethical guidelines, disclosure norms, and educational strategies that equip researchers to use AI responsibly. As technology evolves, so too must the standards of scientific publishing, ensuring that innovation enhances—rather than compromises—the integrity of neurosurgical literature.

References

1) Fauziah RR, Puspita AMI, Yuliana I, Ummah FS, Mufarochah S, Ramadhani E. Artificial intelligence in academic writing: Enhancing or replacing human expertise? *J Clin Neurosci*. 2025 Mar 19:111193. doi: 10.1016/j.jocn.2025.111193. Epub ahead of print. PMID: 40113532.

2) Schneider DM, Mishra A, Gluski J, Shah H, Ward M, Brown ED, Sciubba DM, Lo SL. Prevalence of Artificial Intelligence-Generated Text in Neurosurgical Publications: Implications for Academic Integrity and Ethical Authorship. *Cureus*. 2025 Feb 16;17(2):e79086. doi: 10.7759/cureus.79086. PMID: 40109787; PMCID: PMC11920854.

3) Akgun MY, Savasci M, Gunerbuyuk C, Gunara SO, Oktenoglu T, Ozer AF, Ates O. Battle of the authors: Comparing neurosurgery articles written by humans and AI. *J Clin Neurosci*. 2025 Feb 25;135:111152. doi: 10.1016/j.jocn.2025.111152. Epub ahead of print. PMID: 40010170.

4) Nabata KJ, AlShehri Y, Mashat A, Wiseman SM. Evaluating human ability to distinguish between ChatGPT-generated and original scientific abstracts. *Updates Surg*. 2025 Jan 24. doi: 10.1007/s13304-025-02106-3. Epub ahead of print. PMID: 39853655.

From:

<https://neurosurgerywiki.com/wiki/> - **Neurosurgery Wiki**

Permanent link:

<https://neurosurgerywiki.com/wiki/doku.php?id=xxxxx>

Last update: **2025/04/30 18:00**

