



Artificial intelligence as author: Can scientific reviewers recognize GPT-4o-generated manuscripts?

Ahmet Öztürk, MD^a, Anılcan Tahsin Karahan, MD^a, Serkan Günay, MD^a, Abdul Samed Erdal, MD^b, Seval Komut, MD^a, Erdal Komut, MD^c, Yavuz Yiğit, MD^{d,e,*}

^a Hitit University Erol Olçok Education and Research Hospital, Department of Emergency Medicine, Çorum, Turkey

^b Hitit University Erol Olçok Education and Research Hospital, Department of Pharmacy, Çorum, Turkey

^c Hitit University Erol Olçok Education and Research Hospital, Department of Radiology, Çorum, Turkey

^d Hamad Medical Corporation, Hamad General Hospital, Department of Emergency Medicine, Doha, Qatar

^e Blizard Institute, Queen Mary University, Department of Emergency Medicine, London, UK

ARTICLE INFO

Article history:

Received 27 April 2025

Received in revised form 27 May 2025

Accepted 14 July 2025

Available online xxxx

Keywords:

ChatGPT

Artificial intelligence

Journal

Editor

Reviewer

ABSTRACT

Introduction: Chat Generative Pre-Trained Transformer (ChatGPT) is a natural language processing model. It can be argued that ChatGPT has recently begun to assume the role of a technological assistant capable of supporting academics in the process of scientific writing. ChatGPT may contribute to the spread of incorrect or incomplete information within academic literature, leading to conceptual confusion and potential academic misconduct. The aim of this study is to determine whether a scientific article entirely generated by an AI application such as ChatGPT can be detected by an academic journal editor or peer reviewer.

Methods: This study was conducted between November 1, 2024, and December 1, 2024. GPT-4o, was utilized in this study. ChatGPT was instructed to write a scientific article focused on predicting mortality and return of spontaneous circulation (ROSC) in OHCA cases. The manuscript written by ChatGPT-4o was sent to 14 different reviewers who had previously served as reviewers or editors. The reviewers were asked to evaluate the manuscript as if they were an SCI-E journal editor or peer reviewer. The reviewers were informed that the article had been written by ChatGPT and were asked whether they had identified this during their review.

Results: Among the reviewers, 42.9 % ($n = 6$) decided to reject the manuscript at the editorial stage, whereas another 42.9 % ($n = 6$) opted to forward it to a peer reviewer. During the peer review stage, 42.9 % ($n = 6$) of the reviewers recommended rejection, while 28.6 % ($n = 4$) suggested major revisions. 78.6 % ($n = 11$) of the reviewers did not realize that the manuscript had been generated by an artificial intelligence model.

Conclusion: The findings of our study highlight the necessity for journal editors and peer reviewers to be well-informed about ChatGPT and to develop systems capable of identifying whether a manuscript has been written by a human or generated by artificial intelligence.

© 2025 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Chat Generative Pre-Trained Transformer (ChatGPT) is a natural language processing model with approximately 175 billion parameters that uses deep learning algorithms trained on large datasets to generate human-like responses to user prompts. As a general-purpose dialogue tool, ChatGPT is designed to respond to a wide range of topics, making it potentially useful for applications such as customer service, chatbots, and many others [1]. Moreover, it can be argued that ChatGPT has recently begun to assume the role of a technological assistant capable of

supporting academics in the process of scientific writing. Its ability to automate critical stages of scientific manuscript preparation—such as text generation and data analysis—can be considered a facilitator, and perhaps even a promising development, for researchers. This is made possible by its training on an extensive body of text, allowing it to generate human-like textual responses [2].

In recent years, it has been demonstrated that artificial intelligence (AI) can produce coherent language, and distinguishing AI-generated sentences from those written by humans has become increasingly difficult. In 2022, Nature reported that scientists were using conversational agents as research assistants to help organize their thoughts, receive feedback on their work, write code, and even summarize research literature [2]. In academic papers, ChatGPT is often used during the drafting phase. It can assist researchers in creating a preliminary draft of a

* Corresponding author at: Hamad Medical Corporation, Hamad General Hospital, Doha, Qatar.

E-mail address: yyigit@hamad.qa (Y. Yiğit).

manuscript, revising the generated content, and designing summaries, thereby contributing to a more efficient management of the writing process. Although these drafts can save researchers time and energy, they may not always meet expectations [3].

Despite the advantages it provides, the widespread use of ChatGPT may negatively impact the development of critical and independent thinking skills. It may also introduce incorrect or incomplete information into academic literature, increasing the risk of conceptual confusion and academic misconduct. In particular, concerns include plagiarism, fabricated data, ethical violations, and the inclusion of non-scientific or inaccurate content in academic manuscripts [4]. While current literature shows growing interest in the use of AI tools during the writing process of scientific articles, one of the most significant risks appears to be the generation of fabricated data. Although there are tools designed to monitor the use of AI in manuscript preparation, no definitive mechanism exists to detect whether the data itself has been produced by AI [5]. Due to these risks, even prestigious journals such as *The Lancet* have stated that the responsibility for any use of AI within a manuscript lies entirely with the authors. This highlights the difficulty that even well-established, high-impact journals face in detecting AI usage within scientific publications [6].

The aim of this study is to determine whether a scientific article entirely generated by an AI application such as ChatGPT can be detected by an academic journal editor or peer reviewer. A secondary aim is to assess whether such a manuscript would be accepted by any journal indexed in the Science Citation Index-Expanded (SCI-E).

2. Methods

This study was conducted between November 1, 2024, and December 1, 2024. Since no real patient data were used, ethics committee approval was not required. One of the most up-to-date models of ChatGPT, namely Generative Pre-Trained Transformer 4 omni (GPT-4o), was utilized in this study. Initially, GPT-4o was asked to generate a dataset based on the assumption that it reflected out-of-hospital cardiac arrest (OHCA) cases.

Specifically, GPT-4o was initially instructed to generate a synthetic dataset simulating OHCA cases over a five-year period, guided by scientific literature. The dataset was to include approximately 1000 patients and key variables such as age, sex, cause of arrest, whether the arrest was witnessed, time to initiation of CPR, initial rhythm, return of spontaneous circulation (ROSC) status, defibrillation use, and blood gas parameters. Following dataset generation, GPT-4o asked to perform statistical analyses to identify predictors of ROSC. Based on the analysis results, GPT-4o was then tasked with drafting a retrospective study manuscript reflecting data from the simulated period between 2018 and 2023, as if conducted at Çorum University Hospital. The manuscript was required to comply with CONSORT guidelines, span approximately 1500 words, and include 20 relevant references. As part of the writing process, GPT-4o was asked to conduct a review of the academic literature to ensure appropriate and accurate citations. An initial draft was produced and subsequently reviewed by two independent emergency medicine specialists, who evaluated its adherence to reporting standards. Following the initial draft, two independent researchers reviewed the manuscript and identified sections that did not comply with the CONSORT checklist. The model was then prompted to revise those sections accordingly. In total, four rounds of revision were provided by the researchers to ChatGPT to ensure the accuracy of these technical corrections. Notably, no researcher contributed to either the dataset or the content of the manuscript during its preparation. The abstract was structured according to conventional academic formatting—Introduction, Materials and Methods, Results, Conclusion—and was limited to approximately 250 words. The final version of the manuscript is provided in Supplementary File 1.

In the second phase of the study, the manuscript written by ChatGPT-4o was sent to 14 different reviewers who had previously

served as reviewers or editors for SCI-E indexed journals and had an H-index of 5 or above according to Web of Science. The evaluators who participated in the study were not native English speakers; however, all of them had a good command of the English language. Reviewers were informed in advance that the manuscript they would be assessing may have been generated by an AI model, and they were invited to participate in the study voluntarily. Informed consent was obtained from all reviewers prior to their participation. The manuscript was shared via Google Docs. During this process, no personal identifying information was requested from the reviewers. They were asked to evaluate the manuscript as if they were serving as SCI-E journal editors or peer reviewers. After the evaluations were completed, reviewers were asked whether they had identified the manuscript as AI-generated. Those who responded affirmatively were invited to explain how they made this determination. All responses were recorded and subsequently analyzed statistically.

2.1. Statistical analysis

Statistical analyses were conducted using SPSS version 25 software. Descriptive analyses were presented as numbers and percentages.

3. Results

In our study, the manuscript generated by GPT-4o was evaluated by a total of 14 reviewers, both from the perspective of a journal editor and as a peer reviewer for an SCI-E indexed journal.

Among the reviewers, 42.9 % ($n = 6$) decided to reject the manuscript at the editorial stage, whereas another 42.9 % ($n = 6$) opted to forward it to a peer reviewer. During the peer review stage, 42.9 % ($n = 6$) of the reviewers recommended rejection, while 28.6 % ($n = 4$) suggested major revisions. At the editorial stage, 14.3 % ($n = 2$) of reviewers accepted the manuscript. As peer reviewers, 28.6 % ($n = 4$) of the evaluators recommended acceptance after minor revisions. Detailed results are presented in Table 1.

Additionally, 78.6 % ($n = 11$) of the reviewers did not realize that the manuscript had been generated by an artificial intelligence model. In contrast, 21.4 % ($n = 3$) correctly identified the manuscript as AI-generated. These reviewers identified features such as standardized sentence structures, templated phrasing, a superficial discussion section, and repetitive language as indicators of AI-generated content. Further details are provided in Table 2.

4. Discussion

Only 3 out of the 14 reviewers participating in our study were able to recognize that the manuscript had been written by an artificial intelligence model. Among those asked to evaluate the manuscript from an editorial perspective, 2 reviewers deemed it acceptable at the editorial stage. Meanwhile, 4 of the 14 reviewers who assessed the manuscript as peer reviewers decided that it could be accepted following minor revisions.

Table 1

Evaluation of a manuscript generated by GPT-4o as an SCI-E journal editor or reviewer

Evaluators' response	Evaluation as a SCI-E journal editor (n, %)
Accept	2 (14.3)
Reject	6 (42.9)
Send to review process	6 (42.9)
	Evaluation as a SCI-E journal reviewer (n, %)
Major revision	4 (28.6)
Acceptance after minor revision	4 (28.6)
Accept	0 (0)
Reject	6 (42.9)

GPT-4o: Generative Pre-Trained Transformer 4 omni, SCI-E: Science Citation Index-Expanded.

Table 2
Comments from reviewers who recognized that the manuscript generated by GPT-4o was produced by an artificial intelligence system

Evaluator	Comment
1	Yes, I noticed it. My justifications are as follows: Language Structure and Fluency: • The article generally uses a fairly proper language, but it carries a somewhat artificial “template-like” tone. However, some sentences deviate from natural flow and rely too heavily on technical jargon. General Formatting: • The data, statistics, and tables are presented in a rather mechanical manner. Sections that require more analysis and interpretation—such as the discussion—lack depth. Lack of Originality: • The study's contribution to literature is limited. Although the topic is important, in terms of “novelty” or originality, the structure seems to reflect an AI's capacity to “summarize and interpret existing literature” rather than offering unique insights. Critique and Discussion Section: • The discussion section is quite superficial. The study's limitations and its contribution to literature seem to be drawn directly from existing sources. Repetitions: • Some ideas are repeated across different sections of the article. This might stem from the AI's tendency to express the same data in varied phrasing across separate segments.
2	The numerical expressions in the statistical analysis section under the “Results” heading are not written in an academic manner (e.g., instead of “25 (45 %)” , expressions such as “45 %1” are used...). The p-values and text formatting are also striking. If the text were to be humanized, it would be very difficult to understand, and suspicions would go beyond mere doubt.
3	The use of standard discussion sentence patterns, lack of in-depth analysis in the discussion, incorrect citation of references, and failure to provide additional findings beyond similar studies.

GPT-4o: Generative Pre-Trained Transformer 4 omni, AI: Artificial intelligence.

Given the recent rise in the academic use of artificial intelligence tools—particularly ChatGPT—journal editors and peer reviewers must become increasingly vigilant about whether AI tools have been used during manuscript preparation. Editors and reviewers should remain aware that they may be required to distinguish between manuscripts written based on genuine data and those involving fabricated or misleading content. An article published by *The Guardian* in 2013 reported that science journalist John Bohannon submitted a fabricated manuscript to 304 open-access journals worldwide. The article, which contained falsified data, was accepted by 157 journals, including those published by major publishing houses such as Sage, Elsevier, and Wolters Kluwer [7]. Considering that this occurred at a time when AI tools were far less developed and widely used, the potential acceptance of fabricated manuscripts generated by today's sophisticated AI technologies represents a serious academic threat. The responsibility for distinguishing whether a manuscript was written by a human, or an AI model will often fall on the reviewers or editors, posing a significant challenge.

Indeed, in our study, 78.6 % of the reviewers failed to recognize that the manuscript was generated entirely by ChatGPT. This finding suggests that journal editors and reviewers may not yet be adequately prepared to evaluate AI-generated scientific content. Furthermore, since the reviewers were aware that they were participating in a research study during the consent process, they may have exhibited more meticulous and attentive behavior. If the potential Hawthorne effect is disregarded, it is possible that some reviewers might have been more likely to misidentify whether the manuscript was AI-generated. This suggests that the actual impact of our study may be more pronounced than the results indicate. Derga et al., in a recent study, highlighted this issue and recommended that reviewers, editors, and publishers familiarize themselves with ChatGPT in order to understand its capabilities and limitations. They argued that this would help stakeholders monitor the development of such tools and identify potential pitfalls [2]. Similarly, we believe that experience with AI applications can support the identification of AI-generated manuscripts based on flow, content, and sentence structure. However, as AI continues to advance and approaches a level of sophistication comparable to human writing, even experienced reviewers may struggle to detect AI-generated content.

For this reason, requiring authors to declare that their manuscript has not been generated by AI may help academic journals prevent the publication of deceptive content. Moreover, software tools have been

developed to assist in detecting whether a manuscript was written using AI. In one such study, the “Gotcha GPT” application was used to detect AI involvement in manuscript writing. The study reported that the tool was successful in identifying AI authorship with an accuracy rate of 97–99 % [5]. However, these detection tools also present limitations. The ability of AI models to generate text patterns that closely resemble human writing may lead to suspicions about manuscripts genuinely written by humans. Additionally, most detection tools focus on identifying specific textual patterns [5]. While some software might detect the linguistic features indicative of AI authorship, they may fail to recognize whether the underlying data was also generated by AI. This may, in turn, complicate the detection of fabricated AI-generated publications and facilitate their dissemination.

While the rapid advancement of AI presents growing risks—such as the ability to fabricate entire datasets and write plausible manuscripts—it also offers valuable support. Used appropriately, AI can serve as an efficient tool to assist researchers with formatting, summarization, and translation [3]. In particular, AI tools may assist with manuscript formatting, figure generation, and translation. However, they should never be actively involved in data generation or the complete writing of a manuscript. In one study, Hegde et al. prepared a case report by placing sentences generated by ChatGPT beneath their own. They aimed to highlight the differences between human and AI-generated writing. Although ChatGPT was used in the manuscript, the accuracy and reliability of its content and sources could not be verified [8]. This case illustrates that despite its capacity for coherent text generation, ChatGPT may still pose a risk of introducing inaccurate or unverifiable information into academic literature. Therefore, it is essential that reviewers possess sufficient awareness and practical experience to identify potentially questionable contributions made by ChatGPT in scientific manuscripts.

Undoubtedly, AI-generated or AI-assisted manuscripts increasingly present both benefits and risks for academia and scientific literature. As a preliminary measure in the technical context, scientific platforms might consider requiring comprehensive documentation of each phase of the research process—including datasets—for all submitted manuscripts. However, the rapid advancement of AI in generating datasets, synthetic patients, and disease models is likely to exacerbate this issue in the short term. To partially legitimize potential ambiguities surrounding AI use, academic journals could implement structured declarations during the submission process,

combining multiple-choice and open-ended questions to assess the extent and nature of AI involvement. In cases where AI-generated content is detected beyond the declared scope, journal-specific restrictions could be imposed on the authors. Nevertheless, none of these technical precautions are likely to be as effective as peer reviewers who are experienced and well-versed in AI and AI-driven research, or the deployment of specialized detection software designed for such purposes. It is evident that this issue will become a subject of increasingly intense debate in the near future.

4.1. Limitations

Our study has several limitations. The first is that some of the reviewers had no prior experience with AI tools such as ChatGPT and therefore may not have been familiar with the typical linguistic patterns of AI-generated content. Another limitation is that all interactions with ChatGPT during the study were conducted in English. Therefore, the data obtained in this study do not provide insights into interactions conducted in other languages. Another important limitation is the possibility that reviewers did not approach the manuscript with the same level of seriousness as they would a submission from an actual academic journal. Additionally, the number of reviewers may be considered relatively small, as no formal sample size calculation was performed in this study. Since ChatGPT does not follow a fixed algorithm for manuscript writing, its output varied based on the guidance it received during the generation process, which may have influenced the reviewers' decisions. Finally, during our study, reviewers who identified that the manuscript was generated by AI were asked to explain the factors that led to their recognition. However, no specific questions were directed to those reviewers who failed to detect the AI-generated nature of the manuscript regarding their perspectives in this context. Therefore, our study does not include findings related to this aspect.

5. Conclusion

Our findings underscore the urgent need for journal editors and peer reviewers to remain informed about the capabilities and limitations of tools like ChatGPT. Developing systems to detect AI-generated manuscripts is critical to maintaining trust in the peer-review process and preserving the reliability of scientific literature.

While ChatGPT can serve as a supportive tool in academic writing, its unregulated use poses significant risks—including data fabrication, authorship ambiguity, and ethical concerns. The role of AI in scientific publishing continues to spark debate, particularly regarding the boundaries of acceptable use.

To address these challenges, editorial boards and academic institutions must establish clear, enforceable guidelines to ensure that the integration of AI tools strengthens, rather than undermines, the integrity of scholarly communication.

CRedit authorship contribution statement

Ahmet Öztürk: Supervision, Data curation, Writing – review & editing, Methodology, Conceptualization, Writing – original draft, Investigation. **Anılcan Tahsin Karahan:** Resources, Conceptualization, Writing – review & editing, Formal analysis, Visualization, Data curation.

Serkan Günay: Data curation, Visualization, Formal analysis. **Abdul Samed Erdal:** Investigation, Writing – review & editing, Methodology. **Seval Komut:** Writing – review & editing, Conceptualization, Supervision, Methodology. **Erdal Komut:** Writing – review & editing, Supervision, Investigation. **Yavuz Yiğit:** Project administration, Methodology, Writing – review & editing, Conceptualization.

Ethical approval

This study did not involve human subjects in a clinical or biomedical context and did not include the use of any real patient data. The participants were academic journal reviewers and editors acting in their professional capacity, and no sensitive or personally identifiable information was collected. All participants were informed in advance that the manuscript they were asked to evaluate might have been generated by an artificial intelligence model. Participation was entirely voluntary, and informed consent was obtained electronically prior to their involvement. Given the absence of patient data and the minimal-risk nature of the study, formal ethical approval was not required.

Financial support

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

None to declare.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ajem.2025.07.034>.

References

- [1] Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, et al. How does ChatGPT perform on the United States medical licensing examination (USMLE)? The implications of large language models for medical education and knowledge assessment. *JMIR Med Educ*. 2023;9:e45312. doi:10.2196/45312.
- [2] Dergaa I, Chamari K, Zmijewski P, Ben Saad H. From human writing to artificial intelligence generated text: examining the prospects and potential threats of ChatGPT in academic writing. *Biol Sport*. 2023;40(2):615–22. doi:10.5114/biolSport.2023.125623.
- [3] Ahaley SS, Pandey A, Juneja SK, Gupta TS, Vijayakumar S. ChatGPT in medical writing: a game-changer or a gimmick? *Perspect Clin Res*. 2024;15(4):165–71. doi:10.4103/picr.picr_167_23.
- [4] Doru B, Maier C, Busse JS, Lücke T, Schönhoff J, Enax-Krumova E, et al. Detecting artificial intelligence-generated versus human-written medical student essays: Semirandomized controlled study. *JMIR Med Educ*. 2025;11:e62779. doi:10.2196/62779.
- [5] Gralha JG, Pimentel AS. Gotcha GPT: ensuring the integrity in academic writing. *J Chem Inf Model*. 2024;64(21):8091–7. doi:10.1021/acs.jcim.4c01203.
- [6] Armitage R. Generative AI in medical writing: co-author or tool? *Br J Gen Pract*. 2024;74(740):126–7. doi:10.3399/bjgp24X736605.
- [7] Shaw C. Hundreds of open access journals accept fake science paper. *The Guardian*; 2013 Oct 4.
- [8] Hegde A, Srinivasan S, Menon G. Extraventricular Neurocytoma of the posterior Fossa: a case report written by ChatGPT. *Cureus*. 2023;15(3):e35850. doi:10.7759/cureus.35850.