

# ChatGPT and the Future of Journal Reviews: A Feasibility Study

Som Biswas\*, Dushyant Dobaria, and Harris L. Cohen

*Le Bonheur Children's Hospital, The University of Tennessee Health Science Center, Memphis, TN, USA*

The increasing volume of research submissions to academic journals poses a significant challenge for traditional peer-review processes. To address this issue, this study explores the potential of employing ChatGPT, an advanced large language model (LLM), developed by OpenAI, as an artificial intelligence (AI) reviewer for academic journals. By leveraging the vast knowledge and natural language processing capabilities of ChatGPT, we hypothesize it may be possible to enhance the efficiency, consistency, and quality of the peer-review process. This research investigated key aspects of integrating ChatGPT into the journal review workflow. We compared the critical analysis of ChatGPT, acting as an AI reviewer, to human reviews for a single published article. Our methodological framework involved subjecting ChatGPT to an intricate examination, wherein its evaluative acumen was juxtaposed against human-authored reviews of a singular published article. As this is a feasibility study, one article was reviewed, which was a case report on scurvy. The entire article was used as an input into ChatGPT and commanded it to "Please perform a review of the following article and give points for revision." Since this was a case report with a limited word count the entire article could fit in one chat box. The output by ChatGPT was then compared with the comments by human reviewers. Key performance metrics, including precision and overall agreement, were judiciously and subjectively measured to portray the efficacy of ChatGPT as an AI reviewer in comparison to its human counterparts. The outcomes of this rigorous analysis unveiled compelling evidence regarding ChatGPT's performance as an AI reviewer. We demonstrated that ChatGPT's critical analyses aligned with those of human reviewers, as evidenced by the inter-rater agreement. Notably, ChatGPT exhibited commendable capability in identifying methodological flaws, articulating insightful feedback on theoretical frameworks, and gauging the overall contribution of the articles to their respective fields. While the integration of ChatGPT showcased immense promise, certain challenges and caveats surfaced. For example, ambiguities might present with complex research articles, leading to nuanced discrepancies between AI and human reviews. Also figures and images cannot be reviewed by ChatGPT. Lengthy articles need to be reviewed in parts by ChatGPT as the entire article will not fit in one chat/response. The benefits consist of reduction in time needed by journals to review the articles submitted to them, as well as an AI assistant to give a different perspective about the research papers other than the human reviewers. In conclusion, this research contributes a groundbreaking foundation for incorporating ChatGPT into the pantheon of journal reviewers. The delineated guidelines distill key insights into operationalizing ChatGPT as a proficient reviewer within academic journal frameworks, paving the way for a more efficient and insightful review process.

\*To whom all correspondence should be addressed: Som Biswas, Pediatric Radiology, Le Bonheur Children's Hospital, Memphis, TN; Email: sbiswas8@uthsc.edu; ORCID: 0000-0002-4038-5844.

Abbreviations: LLM, large language model; AI, artificial intelligence

Keywords: chatGPT, review, journal reviews

## INTRODUCTION

The peer-review process is a cornerstone of scholarly publishing, ensuring the quality and integrity of scientific research. However, with the exponential growth of academic submissions, the traditional peer-review system is facing significant challenges in terms of efficiency and scalability [1-3]. One quantitative analysis suggested that the volume of scientific manuscript submissions to journals doubles every 15 years [4]. As the scientific community strives to address these issues, artificial intelligence (AI) technologies have emerged as a potential solution to augment the traditional review process. Among these technologies, ChatGPT, a state-of-the-art large language model (LLM) developed by OpenAI ChatGPT [5], holds promise as an AI reviewer for academic journals.

ChatGPT is an advanced LLM trained on an extensive amount of text from diverse sources, specifically around 570GB of datasets, including web pages, books, and other sources, enabling it to generate coherent and contextually relevant responses. With its ability to understand and process natural language, ChatGPT has demonstrated proficiency in various tasks, including text completion, question-answering, and language translation [6,7]. It has also been used in conducting a literature review [8].

The objective of this study is to investigate the viability of ChatGPT as an AI reviewer for academic journals. By leveraging its vast knowledge and natural language processing capabilities, ChatGPT has the potential to streamline and augment the review process. This could lead to benefits such as faster manuscript evaluation, increased consistency in feedback, and improved handling of the growing volume of submissions. However, it is crucial to assess the model's performance, identify limitations, and address ethical concerns before implementing it in real-world scenarios.

In this article, we present a pilot analysis of the integration of ChatGPT into the journal review workflow. We evaluate ChatGPT's performance as an AI reviewer, examining its ability to assess scientific methodology, evaluate research novelty and significance, and provide constructive feedback to authors. We do this by using ChatGPT to assess the initial manuscript submitted to a peer-reviewed journal for publication. The review produced by ChatGPT was then compared with the original human reviews that were sent by the journal to the author.<sup>1</sup> We also discuss the potential advantages of using ChatGPT as a standalone reviewer or potential review assistant. Beyond its potential benefits, we also ac-

knowledge the challenges associated with implementing ChatGPT as an AI reviewer. These include addressing biases inherent in the model, ensuring transparency and explainability, and defining the role of human oversight in the review process. By examining both the opportunities and limitations, we aim to provide insights and guidelines for researchers, publishers, and AI developers interested in adopting ChatGPT as an AI reviewer.

Ultimately, this research contributes to the ongoing discourse on the future of peer review, exploring the potential of AI technologies to augment and optimize the scholarly publishing ecosystem. By embracing innovative approaches while upholding the core principles of rigorous evaluation and intellectual contribution, we can pave the way for a more efficient, inclusive, and impactful scientific community. However, a balance must be struck between the potential of AI and the importance of human judgment in the peer-review process.

## EVALUATION OF THE PERFORMANCE OF ChatGPT

### Methodology

The draft of the published article by the authors titled "Atypical Imaging Findings of Scurvy in A Malnourished Child"<sup>2</sup> was used to evaluate performance of ChatGPT. Author guidelines specific to the journal that had published the article were entered in ChatGPT initially, to fine-tune the review process. These guidelines are available on most journal websites and were copied and pasted into ChatGPT and the command was simultaneously entered to review the following manuscript as per these author guidelines.

Many research papers are long and entire article cannot be fit in one dialogue box. For the purpose of a feasibility study, a small article (case report on scurvy) was used that fit in one dialogue box. However, for lengthy complex articles they need to be broken into parts and then each part needs to be put separately in the dialogue box, one at a time. This entire article was then copied and pasted into the ChatGPT input box and in the same box ChatGPT was commanded to "Please perform a review of the following article and give points for revision." Since this was a case report with a limited word count, the entire article could fit in one chat box. Then the output (the review provided by ChatGPT) was copied into a Word file and was compared with human reviewers' comments specifically for content, coherence, and relevance. As this is a feasibility study only one article that is a case report

<sup>1</sup>The authors have permission to reprint the anonymous reviewer comments.

<sup>2</sup>This article can be read here: Biswas S, Miller S, Cohen HL. Scurvy in A Malnourished Child: Atypical Imaging Findings. J Radiol Case Rep. 2022 Sep 1;16(9):11-15. doi: 10.3941/jrcr.v16i9.4545.

on scurvy was used as a sample. Also, as only one article was used, a qualitative analysis was performed by two human authors who cross-reviewed the results. Only qualitative analysis was done and no quantitative scoring or rating was used. Further research with multiple papers of various types and lengths could be performed in the future.

The following are the outputs generated by ChatGPT in its review process:

#### *Review by ChatGPT of this article:*

**General Comments:** The article titled “Atypical Imaging Findings of Scurvy in a Malnourished Child” presents a case report discussing the diagnosis and imaging characteristics of scurvy. While the article addresses an important topic, several areas require improvement to enhance the overall quality and clarity of the manuscript.

**Abstract:** The abstract provides a brief overview of the article; however, it lacks essential details. It would benefit from including specific information, such as the age of the patient, a brief mention of the treatment regimen, and the significance of the case in the context of the existing literature.

**Introduction:** The introduction effectively highlights the rarity of scurvy in developed countries, but it lacks a clear statement of the objective or research question. It would be beneficial to provide a concise statement of the study’s purpose and its relevance to the medical community.

**Case Report:** The case report provides a detailed description of the patient’s presentation and diagnostic journey. However, it would be helpful to include more information about the patient’s demographic characteristics, as well as additional relevant medical history. Furthermore, the description of the radiographic and MRI findings could be enhanced by including specific details such as image sequences, locations, and any measurements taken. This would improve the reproducibility of the study and aid in comparing the findings to existing literature.

**Discussion:** The discussion section is informative but lacks depth in certain areas. While it briefly touches upon the etiology, demographics, clinical findings, and treatment of scurvy, it does not delve into the underlying mechanisms of the disease or provide a comprehensive analysis of the existing literature. The authors should expand on these aspects and critically analyze the findings in light of previous studies, highlighting any novel or conflicting observations.

The discussion of imaging findings is informative, but it would benefit from a more comprehensive analysis. While the authors mention the classical radiographic features of scurvy, they fail to provide a clear explanation

of why the reported case exhibited atypical findings. Further exploration of potential explanations or hypotheses would strengthen the discussion.

The differential diagnoses section is brief and lacks depth. It would be valuable to provide a more thorough discussion of the differential diagnoses considered, including other nutritional deficiencies or metabolic disorders that could present with similar imaging findings.

**Conclusion:** The conclusion is concise but fails to emphasize the key takeaways from the study. It should highlight the significance of the atypical imaging findings, the importance of considering scurvy in the differential diagnosis, and the implications for clinical practice.

**Overall Structure and Language:** The article would benefit from a clearer structure and improved flow. The language used is generally clear, but there are instances where the writing could be refined for better readability and organization.

In summary, while the article provides some valuable insights into atypical imaging findings of scurvy, it requires significant improvements. Strengthening the abstract, enhancing the case report section, conducting a more thorough literature review, expanding the discussion, and refining the conclusion would significantly enhance the manuscript’s quality and impact. Additionally, attention to the overall structure and language would improve the overall readability and flow of the article.

#### *Review by human authors before publication of this article:*

**Reviewer 1:** Recommended correction of grammatical errors.

**Reviewer 2:** The author discusses the occurrence of a rare disease, scurvy, in children. It is noteworthy that vitamin C deficiency is uncommon in developed countries, including the case presented in the manuscript. However, there is a reported case of “Vitamin C deficiency in a picky eater child” documented by Fickrey et al. in the Medical Journal of Malaysia in 2021 (PMID: 34508390). Furthermore, several studies have documented cases of scurvy in children with neurological disorders and/or unusual dietary patterns, such as the ketogenic diet.

To enhance the clarity of the manuscript, please provide clarification regarding the presence of any neurological impairment in the child with Noonan syndrome.

Additionally, it is recommended to include the following references in the discussion section:

Skeletal Radiol 2015: “MRI findings in pediatric patients with scurvy.” doi: 10.1007/s00256-014-1962-y.

Pediatr Radiol 2013: “Scurvy in an autistic child: MRI findings.” doi: 10.1007/s00247-013-2688-z.

**Reviewer 3:** While the manuscript is well-written in proficient English and possesses substantial educational value, I would like to provide some constructive feed-

back:

Page 2, line 6: Change “No changes in myositis were seen” to “No changes of myositis were observed.”

Page 2, line 14: Please include the normal ranges when stating “Ascorbic acid levels were determined to be low.”

Page 2, line 20: The phrase “or collagen formation, bone matrix, collagen, tendon, and ligament synthesis” repeats the mention of collagen formation. Please revise for clarity.

Page 4, line 10: Amend “These physiological changes need” to “These physiological changes require.”

Page 5, line 9: It is advisable to tone down the statement “This makes this case report very unique.” Title: Modify the title to “In a Malnourished Child.” Abstract: Omit the references mentioned in the abstract.

Figure legend 1: The description seems contradictory with “...metaphyseal lucent band in the proximal femur. It appears radiodense in the metaphysis...” Clarify the terms “radiolucent” and “radiodense.”

Figure legends: Enhance the descriptive quality of the figure legends.

Case report: When discussing imaging features, refer to the corresponding figures.

Figures: Increase the size of the figures to enable better visibility of details. It is recommended to add magnifications. MRI sequences: Include additional MRI sequences to augment the presentation. Differential diagnosis (DD) table: Revise the DD table according to the author guidelines, with columns for modalities and rows for entities/DDs.

Tables: Rectify the typographical errors present in the tables.

Question 5: Correct the name from “Summerfield” to “Trummerfeld.”

Based on these suggestions, I recommend the publication of this manuscript once these revisions have been implemented.

**Editor’s comment:** The editor believes that despite the critiques raised by the reviewers, the article is suitable for publication in the journal. However, it is necessary to incorporate the suggested revisions and address the reviewers’ concerns before proceeding with the publication.

Please revise your manuscript accordingly, taking into account the feedback from the reviewers and editor. Once you have made the necessary revisions, you can submit the revised version for final consideration.

## POTENTIAL BENEFITS AND CHALLENGES ASSOCIATED WITH INTEGRATING ChatGPT INTO THE EXISTING PEER-REVIEW FRAMEWORK

Integrating ChatGPT into the existing peer-review framework can offer several potential benefits, but it also presents certain challenges. This section explores both these aspects.

### Benefits

**Efficiency and Speed:** ChatGPT can expedite the peer-review process by assisting with initial screenings, providing quick feedback on manuscript suitability, and identifying potential flaws or ethical concerns. This can reduce the burden on human reviewers and speed up the overall publication timeline [9].

**Consistency and Standardization:** ChatGPT can help ensure a more consistent and standardized review process. It can follow predefined criteria and guidelines consistently, reducing subjective biases and variations that can arise among human reviewers [10].

**Language Support:** ChatGPT can aid in overcoming language barriers. ChatGPT has the capability to provide translation services and assist non-native English-speaking authors and reviewers by providing language suggestions or improving clarity in writing [11].

**Accessible Feedback:** ChatGPT can provide detailed feedback and explanations to authors, guiding them on how to improve their manuscripts. This feedback can be available at any time, providing a valuable resource to authors even outside the traditional review process [12].

### Challenges

**Contextual Understanding and Expertise:** ChatGPT might struggle with in-depth comprehension and contextual understanding of highly specialized or nuanced scientific topics. It may not possess the expertise to fully assess the scientific validity or accuracy of complex research findings.

**Subjectivity and Interpretation:** Peer review often involves subjective judgments and interpretations. ChatGPT might lack the ability to understand the broader implications or subtle nuances that human reviewers can identify, potentially missing important considerations.

**Ethical Concerns and Bias:** There is a risk of perpetuating biases or unintentional discrimination in the review process if ChatGPT is trained on biased or unrepresentative datasets. Careful attention must be given to training data selection, model biases, and regular audits to mitigate these issues.

**Lack of Human Element:** Peer review traditionally involves a human connection, enabling nuanced discussions and exchanges of ideas. Integrating ChatGPT might diminish the personal interaction and collaboration between authors and reviewers, potentially impacting the quality of feedback and the iterative refinement of scientific work.



**User Confidence and Acceptance:** There might be initial skepticism or resistance to relying heavily on AI in the peer-review process. Building trust, ensuring transparency, and addressing concerns about accountability and bias will be crucial to gaining acceptance and adoption.

It is important to note that while ChatGPT can augment and support human reviewers, it is not intended to replace them. A balanced approach that combines the strengths of AI with human expertise is key to achieving the best outcomes in the peer-review process.

**Lack of Image Interpretation:** As of now, ChatGPT cannot review figures submitted as part of manuscripts. However, there are emerging AI tools for image analysis that may be able to complement this in the future [13,14]. As comprehension of figures and visual content in scientific research is crucial for conveying complex information, it's important to acknowledge that the limitations of ChatGPT in regard to reviewing these images, submitted as part of manuscripts to the journals, can hinder the complete review of the article. However, there are promising developments in the field of AI-powered image analysis that may soon complement text reviews.

**Guidelines for implementing ChatGPT as a reviewer** for journals can help ensure a standardized and effective integration of AI technology into the peer-review process. To achieve this, several proposed guidelines can be considered. First, clear objectives should be defined, outlining ChatGPT's role and responsibilities. For example, the reviews and comments made by ChatGPT need to be checked by a human, like the editor before sending it out to the authors. Transparency and disclosure are crucial, as authors should be made aware of the AI involvement and understand how ChatGPT functions, its limitations, and its potential impact on the review outcome. For example, authors need to be informed which of the review comments are made by AI and which by human reviewers. Collaboration between ChatGPT and human reviewers should be encouraged to ensure comprehensive assessments. Feedback mechanisms should be established for authors to provide input on ChatGPT-generated reviews, enabling iterative improvements. For example, the author replies can be used by the editor as further training data for ChatGPT to improve itself as a reviewer. Ethical concerns, bias, and diversity must be actively monitored and mitigated. Clear guidelines should be provided to human reviewers on effectively incorporating ChatGPT's feedback. To foster a balanced and effective peer-review process, the successful implementation of ChatGPT as a reviewer requires periodic evaluation, continuous training, user support, and transparency.

## CONCLUSION

In conclusion, the integration of ChatGPT as a reviewer in the journal peer-review process offers both potential benefits and challenges. The benefits include increased efficiency, standardized and consistent evaluations, language support, and accessible feedback. However, challenges arise from ChatGPT's limitations in contextual understanding, potential biases, lack of human interaction, and subjective interpretation. To implement ChatGPT effectively, it is crucial to define clear objectives, ensure transparency and disclosure, train and calibrate the model, promote active collaboration between AI and human reviewers, address ethical concerns and biases, provide clear guidelines, evaluate performance periodically, and invest in continuous training and development.

While ChatGPT can enhance the peer-review process, it should not replace human reviewers entirely. A balanced approach that combines AI technology with human expertise is essential. Maintaining the integrity and quality of the peer-review process requires ongoing efforts to build trust, address biases, promote inclusivity, and foster collaboration between AI and human reviewers. Implementing these guidelines will contribute to an efficient, transparent, and rigorous peer-review system that upholds ethical publishing practices and embraces diversity.

As AI technology continues to advance, ongoing research, evaluation, and refinement of AI-driven peer-review systems will be necessary. By leveraging the strengths of AI while recognizing its limitations, the scientific community can harness the transformative potential of ChatGPT and other AI tools to further enhance the peer-review process and advance the dissemination of high-quality scientific research. Specifically, AI can be used to assist in providing constructive feedback and comments to the authors regarding their submissions.

## REFERENCES

1. Iyengar KP, Jain VK, Ish P. Publication Surge in COVID-19: The Flip Side of the Coin! *J Orthop Traumatol Rehabil*. 2021;13(2):180.
2. Leopold SS. Editorial: increased manuscript submissions prompt journals to make hard choices. *Clin Orthop Relat Res*. 2015 Mar;473(3):753–5.
3. Zuengler J, Carroll H. Reflections on the Steady Increase in Submissions. *Mod Lang J*. 2010;94(4):637–8.
4. Larsen PO, von Ins M. The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index. *Scientometrics*. 2010 Sep;84(3):575–603.
5. <https://openai.com/blog/ChatGPT>. Accessed on August 11, 2023.
6. Lee P, Bubeck S, Petro J. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. *N Engl J Med*. 2023

- Mar;388(13):1233–9.
7. Ayre J, Mac OA, McCaffery KJ, McKay BR, Liu M, Shi Y. New frontiers in health literacy: Using ChatGPT to simplify health information for people in the community. *medRxiv*, 2023-07. <https://doi.org/10.1101/2023.07.24.23292591>.
  8. Ayers JW, Zhu Z, Poliak A, Leas EC, Dredze M, Hogarth M, et al. Evaluating Artificial Intelligence Responses to Public Health Questions. *JAMA Netw Open*. 2023 Jun;6(6):e2317517–2317517.
  9. Hosseini M, Horbach SP. Fighting reviewer fatigue or amplifying bias? Considerations and recommendations for use of ChatGPT and other large language models in scholarly peer review. *Res Integr Peer Rev*. 2023 May;8(1):4.
  10. Al-Mhdawi MK, Qazi A, Alzarrad A, Dacre N, Rahimian F, Buniya MK, et al. Expert Evaluation of ChatGPT Performance for Risk Management Process Based on ISO 31000 Standard. Available at SSRN 4504409.
  11. Kohnke L, Moorhouse BL, Zou D. ChatGPT for Language Teaching and Learning. *RELC J*. 2023:00336882231162868.
  12. Dai W, Lin J, Jin F, Li T, Tsai YS, Gasevic D, Chen G. Can Large LLMs Provide Feedback to Students? A Case Study on ChatGPT.
  13. Cherni MA, Sayadi M, Fnaiech F. AI Tools in Medical Image Analysis: Efficacy of ANN for Oestrogen Receptor Status Assessment in Immunohistochemical Staining of Breast Cancer. *Int J Biomed Eng Technol*. 2013;12(1):60–83.
  14. Trucco E, MacGillivray T, Xu Y, editors. *Computational Retinal Image Analysis: Tools, Applications and Perspectives*. Academic Press; 2019.