Health Affairs
Scholar
EMERGING & GLOBAL HEALTH POLICY

OXFORD

Use of artificial intelligence and the future of peer review

Howard Bauchner[1],* and Frederick P. Rivara[2]

[1]Boston University Chobanian & Avedisian School of Medicine, Visiting Scholars Program, National University of Singapore, 02118, Singapore
[2]Department of Pediatrics, University of Washington, Seattle, WA 98195, United States

*Corresponding author. Boston University Chobanian & Avedisian School of Medicine, Visiting Scholars Program, National University of Singapore, 02118, Singapore. Email: Howard.bauchner@gmail.com

Abstract

Conducting high-quality peer review of scientific manuscripts has become increasingly challenging. The substantial increase in the number of manuscripts, lack of a sufficient number of peer-reviewers, and questions related to effectiveness, fairness, and efficiency, require a different approach. Large-language models, 1 form of artificial intelligence (AI), have emerged as a new approach to help resolve many of the issues facing contemporary medicine and science. We believe AI should be used to assist in the triaging of manuscripts submitted for peer-review publication.

Key words: peer review; scientific communication; AI.

Introduction

The future of peer review of scientific manuscripts should include artificial intelligence (AI). Large-language models, which incorporate in-context learning, could be used to assist editors in triaging manuscripts. Numerous questions have emerged over decades with respect to peer review. First, is it effective —that is, can it detect fabrication, falsification, image manipulation, or adherence with the hundreds of reporting guidelines that have been developed? Second, is it fair, or subject to various forms of bias? Third, is it efficient, or increasingly labor intensive without compensation? Given these challenges, it is inevitable that AI should and will be increasingly used to assist in peer review.

Effectiveness

Most journal editors believe that peer review is essential to ensure high-quality scientific publication.[1] Although peer-reviewers are consultants to editors, and indeed editors generally make the final decision to accept or reject manuscripts informed by the comments from peer review, editors cannot possibly have sufficient expertise to assess every manuscript. Hence, the importance of peer review. But it is well known that rarely can peer-reviewers assess if data have been fabricated or falsified, or if images have been manipulated, which is also a difficult, if not impossible, task for editors. In addition, it is well documented that peer-reviewers often disagree with one another.[2]

Many journals now require that authors indicate if they have adhered to various reporting guidelines—for example, those listed on the EQUATOR website.[3] Yet, there are no data if peer-reviewers (or the editors) assess if authors have adhered to these recommendations, with perhaps the exception of CONSORT (Consolidated Standards of Reporting Trials), the reporting guideline for randomized clinical trials.

Artificial intelligence could very well be more effective at detecting research misconduct and assessing adherence with reporting guidelines than peer-reviewers.

Fairness

There are different types of peer review: open peer review and single- and double-blind peer review. In open peer review the authors and peer-reviewers are known to each other. Peer-reviewers must sign the comments they provide to the authors. One caveat is that some journals that require open peer review still allow confidential comments to the editors. In single-blind peer review the peer-reviewers know the authors (and institutions), but the authors are not provided with the names of the peer-reviewers. In double-blind peer review, theoretically neither the authors nor the peer-reviewers are aware of the identity of the other. Many feel that double-blind peer review is quite difficult to achieve. In laboratory-based science, peer-reviewers are often familiar with other labs that are conducting similar work (and the reason why a peer-reviewer would have been chosen to review). In clinical investigation, particularly randomized clinical trials and meta-analyses, protocols have often been published prior to the completion of studies and included in the reference list, ensuring that peer-reviewers would be familiar with the investigators. In addition, in many fields—for example, oncology, cardiology, and critical care medicine—experts who would serve as peer-reviewers are familiar with the major trials being conducted.

The various forms of peer review have been developed to attempt to deal with well-known biases.[2] For example, some peer-reviewers may favor (or not favor) certain authors and institutions.[2,4] A recent study found that, if peer-reviewers were told that the author of a manuscript had been awarded a Nobel prize, only 23% recommended rejection, compared with 48%

when the author was anonymized, and 65% when the author was relatively unknown.[4] English remains the language of science, and for those who are not native English writers, they may be at a disadvantage when they submit manuscripts. A recent randomized clinical trial reported that 1432 manuscripts submitted to an ecology journal were randomly assigned to single-blind or double-blind review.[5] The authors found that, when reviewers were aware of the authors (single-blind), they gave a more favorable rating from countries with higher English proficiency and higher income. This is one of the largest trials ever conducted comparing double-blind to single-blind review, with findings consistent with what has been known for years, peer-reviewers can be biased.[2,4,5] Artificial intelligence could be developed that is less biased than peer-reviewers.

Efficiency

The number of scientific publications is rising rapidly. It has been estimated to have increased by 47% between 2016 and 2022 from about 1.92 million manuscripts to 2.82 million per year.[6] Most editors are struggling to find enough peer-reviewers for each manuscript. Some journals now suggest, and many now require, that authors list several potential peer-reviewers. As the number of manuscripts has increased, experts who serve as peer-reviewers are increasingly overwhelmed with the number of requests they receive each month. How have journals responded? One way is that many journals have increased the percentage of manuscripts that are rejected by editors without external peer review.

In addition to efficiency, peer review has largely existed as a "free" service. Some journals provide different types of benefits for peer-reviewers—for example, a reduction in open-access fees, a complimentary subscription to the journal, or continuing medical education credits—but most peer-reviewers do not receive any form of financial compensation other than contributing to the community of science. In medicine, the field we know best, physicians are extremely stressed by their daily workload, including devoting time to tasks for which they are not compensated. In addition, generational shifts have led to greater reluctance to extend the work week to accommodate efforts such as peer review. Although AI cannot solve the compensation issue, it can help in peer review, relieving the burden on an already overwhelmed system of review.

The future

There have been a few reports regarding the effectiveness of AI to peer review. For example, Liang and colleagues[7] assessed the agreement between the comments of peer-reviewers and GPT-4 for 3096 submitted manuscripts in 15 of the *Nature* family of journals and 1709 manuscripts submitted to the International Conference on Learning Representations (ICLR). They found generally good agreement, an average overlap of comments of 30.1% for the *Nature* journals, and 35.3% for the ICLR manuscripts. Overlaps between 2 peer-reviewers were similar: 28.6% for the *Nature* journals and 39.2% for ICLR manuscripts. They then asked 308 investigators in the field of AI and computational biology if they found feedback from GPT-4 review to be helpful: 57.4% found it helpful and 82.4% more helpful than feedback from some peer-reviewers. In a recent reported case study in which a single article was compared with the comments from 3 peer-reviewers, the authors concluded: "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."[8] Currently, there are not sufficient data that AI is good enough alone to conduct peer review, but that is most likely to change as systems mature. Artificial intelligence will likely be able to assess adherence with the various reporting guidelines; be less biased with respect to authors, institutions, and language; and perhaps detect fabrication, falsification, or image manipulation.

There are concerns about the use of AI. Most large-language models extract data from thousands if not millions of published works. To do so, AI would add the manuscript being reviewed to the dataset, violating confidentiality with the authors, and placing the authors' work in the public domain. One possible solution would be for large-language models to only use published work from the submitted journal's database, thereby protecting confidentiality. It could also use articles published as open access, since they would be available. In addition, data are needed to determine how good AI is at peer review, specifically with respect to bias and detection of image manipulation, which has emerged as an important issue in contemporary science, as highlighted by recent examples at Stanford and the Dana-Farber Cancer Institute.[9,10]

We envision a future where AI is used to initially scan all submissions and provide a summary of the quality of the manuscript, which will then be reviewed by the editors, prior to a decision to request peer review. This is the inevitable future and will likely debut in some journals within 1 year. Rather than avoiding AI, editors should embrace it. The task then will be to evaluate how good it is and reassure authors that their work has been fairly and appropriately considered by a journal.

Author contributions

Drs. Bauchner and Rivara contributed to all aspects of the manuscript.

Supplementary material

Supplementary material is available at *Health Affairs Scholar* online.

Conflicts of interest

Notes

1. Rennie D. Editorial peer review: its development and rationale. *Peer Rev Health Sci*. 2003;2:1-13.
2. Haffar S, Bazerbachi F, Murad MH. Peer review bias: a critical review. *Mayo Clin Proc*. 2019;94(4):670-676. https://doi.org/10.1016/j.mayocp.2018.09.004

3. EQUATOR Network. Home page. Accessed March 19, 2024. https://www.equator-network.org/

4. Huber J, Inoua S, Kerschbamer R, Konig-Kersting C, Palan S, Smith VL. Nobel and novice: author prominence affects peer review. *Proc Natl Acad Sci USA*. 2022;119:e2205779119. https://doi.org/10.1073/pnas.2205779119

5. Fox CW, Meyer J, Aimé E. Double-blind peer review affects reviewer ratings and editor decisions at an ecology journal. *Funct Ecol*. 2023;37(5):1144-1157. https://doi.org/10.1111/1365-2435.1425

6. Hanson MA, Barreiro PG, Crosetto P. The strain on scientific publishing. arXiv, preprint: not peer reviewed. https://arxiv.org/ftp/arxiv/papers/2309/2309.15884.pdf

7. Liang W, Zhang Y, Cao H, et al. Can large language models provide useful feedback on research papers? A large-scale empirical analysis. arXiv, 2310.01783; October 3, 2023. https://doi.org/10.48550/arXiv.2310.01783

8. Biswas S, Dobaria D, Cohen HL. ChatGPT and the future of journal reviews: a feasibility study. *Yale J Biol Med*. 2023;96:415-420.

9. Freyer FJ, Ryan A. How does bad data slip through? Allegations of research fraud raise questions about "peer review". *The Boston Globe*. Accessed March 20, 2024. https://www.bostonglobe.com/2024/01/28/metro/dana-farber-cancer-institute-retractions/

10. Goldhill O, Moltenni M. "This actually changes everything": altered image in 1999 paper raises potential peril for Stanford president. STAT. November 30, 2022. Accessed March 20, 2024. https://www.statnews.com/2022/11/30/stanford-president-altered-images/