Forging the Way Forward to Inclusive and Responsible Artificial Intelligence in Scholarly Publishing

Sumi Sexton, Chhavi Chauhan, and José E Rodríguez

With the recent Executive Order calling for "removing barriers to American leadership" in artificial intelligence (AI), development of AI and AI-enabled tools in the United States is expected to accelerate. However, in the absence of mandatory checks and balances, it is highly likely that the governance and the quality of output synthesized by generative AI tools may be compromised significantly, and the output may even lead to unintended consequences in the long run.

Like all other domains, the role of AI in scientific publishing is advancing rapidly, such that it is hard to imagine the future processes for writing, reviewing, and editing articles in 25 years, let alone the ways in which processes will change by the end of 2025. Regardless of AI implications on scholarly publishing now or in the distant future, we must ensure that AI is applied in a way that is safe and ethical and helps maintain the rigor and integrity in scholarship.[1-3] Of particular importance is navigating the influence of AI on diversity, equity, inclusion, antiracism, and accessibility (DEIA). Clinical studies have already reported severe (even detrimental) impact on patient populations when AI is widely adopted without validation.[4] This problem is further magnified when AI is trained on limited datasets that are inherently exclusionary and then applied to marginalized groups.[5,6]

Sumi Sexton, MD (https://orcid.org/0000-0001-8574-9237), is Editor in Chief, American Family Physician, and Professor, Department of Family Medicine, Georgetown University School of Medicine. Chhavi Chauhan, PhD, Director of Scientific Outreach, American Society for Investigative Pathology and Founder and President of Samast AI. José E Rodríguez, MD, is Deputy Editor of Family Medicine, and associate vice president of the University of Utah for Health Sciences Workforce Excellence.

Fast forward to the year 2050, when hopefully the publishing landscape includes affordable AI tools developed on robust datasets—empowering efficient editorial workflows, improved searchability, automated accurate language translation, possible alternative formats for both writers and readers, simpler bias detection, and enhanced transparency and accessibility—leading to fair treatment of authors and researchers globally. However, before we can envision this scenario, there are significant challenges to be addressed by multiple stakeholders along the way. These include the potential to create tools that perpetuate existing biases in the literature leading to wider health disparities in underserved groups, lack ethical considerations and cultural nuances, have limited regional access, and are cost prohibitive.[7,8] A key challenge is navigating the shifting political climate which disincentivizes companies from considering some of these factors.

Currently available AI tools are limited in their ability to detect bias as they can only organize language the way it has been seen on the Internet. Because of the biases that exist even in the most objective of places (i.e., scientific journals), the promulgation of AI will only serve to perpetuate bias, instead of eliminating it.[8] For example, medical algorithms that have inappropriately equated race with genetics leading to underuse of lifesaving antihypertensive drug classes and less frequent offering of vaginal birth after cesarean delivery in Black patients have come into question, yet these will continue to surface with use of AI.[9,10] While humans are also not free of bias, it will be nearly impossible to truly eliminate all bias from the AI. AI can be taught to look for bias, but the complex patterns that show bias have yet to be discovered.

In addition to eliminating bias, providing accurate information is critical. The tendency for AI to confabulate/hallucinate is well known, so just imagine if these inaccuracies promote misplaced concepts, falsified data (propagating misinformation/disinformation), and fake papers that negatively impact groups that are already disadvantaged, severely compromising research integrity and eroding trust in scholarship.

CONTINUED

Aside from bias and accuracy, affordability and accessibility of AI tools will be essential to prevent increasing the current (and potentially widening) digital divide. For example, institutions, researchers, communities, cities, and countries across the globe with resources will be able to utilize tools to produce more publications at scale. Yet those in resource-limited settings, including those who may be researching and writing about marginalized populations, may not have access to the same tools, which can lead to less scholarly information being available from these groups. This may easily turn itself into a self-perpetuating vicious cycle of further marginalizing the already marginalized. The role of government infrastructures should also be considered, since some countries provide more access.

So, amid all the existing challenges and emerging chaos, what is the way forward? AI needs humans, like the movement toward DEIA needed humans, specifically with intersectional identities based upon diverse overlapping backgrounds. In the case of AI, these humans would have expertise in both bias and in coding, which is a rare combination, yet represents our best chance to continue to purge the bias that humans introduced into their large language models and other AI tools. Although some companies with the mission to remove bias from healthcare, like Equality AI, closed its doors, there are other larger companies like IBM's watsonx.governance which may have a better chance at addressing bias and promoting ethical practices. Given the lack of guardrails, however, keeping up with the rapid pace of innovation will remain an ongoing challenge.

As we consider AI's potential to exacerbate disparities and to perpetuate misinformation, we must return to the fact that AI is not sentient. It is a computer program. Computer programs do not do what you want them to do, they only do what you tell them to do, within the confines of the desired parameters. Therefore, the governing programs need to seek information from trusted sources and learn to weigh that information more heavily than information from the echo chambers of popular social media amplifying mis/disinformation. Published scholarly works on DEIA must be incorporated into AI algorithms. This is a possible future profession for DEIA leaders in the private sector as government defunding is fully implemented.

Scientific journals would be wise to band together and form partnerships with AI companies to ensure that an AI exists that can provide trusted scientific evidence without the bias and misinformation prevalent on the Internet today.[11] This could mitigate the spread of misinformation; however, it will not eliminate it. The COVID-19 pandemic taught us that even reputable journals can sometimes publish studies that are deeply flawed. WebMD, Doximity, and Medscape have developed AI tools that are already available to physicians, which can provide a credible alternative to those fueled by other (mis)information. However, we must be careful not to trade one type of bias for another, as these collaborations are often heavily subsidized by the pharmaceutical industry.

To summarize, though it may seem daunting to develop, train, and operationalize ethical and responsible AI that is sustainable, scalable, inclusive, and performs optimally and as desired on all needed datasets while meeting all user needs, we must still strive to meet these needs in our own capacities. Only then can we achieve a bright future where responsible AI empowers humans to excel in their domains and enables the betterment of humankind.

References and Links

1. Zielinski C, Winker MA, Aggarwal R, et al. Chatbots, generative AI, and scholarly manuscripts. WAME recommendations on chatbots and generative artificial intelligence in relation to scholarly publications. [accessed March 3, 2025]. https://wame.org/page3.php?id=106.
2. International Committee of Medical Journal Editors. Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals. Updated January 2025. [accessed March 3, 2025]. https://www.icmje.org/icmje-recommendations.pdf.
3. Adams L, Fontaine E, Lin S, Crowell T, Chung VCH, Gonzalez AA, eds. Artificial intelligence in health, health care and biomedical science: an AI code of conduct framework principles and commitments discussion draft. NAM Perspectives. 2024. https://doi.org/10.31478/202403a.
4. Wong A, Otles E, Donnelly JP, et al. External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. JAMA Intern Med. 2021;181:1065–1070. https://doi.org/10.1001/jamainternmed.2021.2626.
5. Larrazabal AJ, Nieto N, Peterson V, Milone DH, Ferrante E. Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis. Proc Natl Acad Sci U S A. 2020;117:12592–12594. https://doi.org/10.1073/pnas.1919012117.
6. Guo LN, Lee MS, Kassamali B, Mita C, Nambudiri VE. Bias in, bias out: underreporting and underrepresentation of diverse skin types in machine learning research for skin cancer detection—a scoping review. J Am Acad Dermatol. 2022;87:157–159. https://doi.org/10.1016/j.jaad.2021.06.884.
7. Garba-Sani Z, Farinacci-Roberts C, Essien A, Yracheta J. A.C.C.E.S.S. AI: a new framework for advancing health equity in health care AI. Health Affairs Forefront. 2024. https://doi.org/10.1377/forefront.20240424.369302.
8. Schrager S, Seehusen DA, Sexton SM, et al. Use of AI in family medicine publications: a joint editorial from journal editors. J Am Board Fam Med. 2025. https://doi.org/10.3122/jabfm.2024.240397R0.
9. Reddick B. Fallacies and dangers of practicing race-based medicine. Am Fam Physician. 2021;104:122–123. https://www.aafp.org/pubs/afp/issues/2021/0800/p122.html.
10. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019;366:447–453 https://doi.org/10.1126/science.aax2342.
11. Ramoni D, Sgura C, Liberale L, Montecucco F, Ioannidis JPA, Carbone F. Artificial intelligence in scientific medical writing: legitimate and deceptive uses and ethical concerns. Eur J Intern Med. 2024;127:31–35. https://doi.org/10.1016/j.ejim.2024.07.012.