

Enacting AI disclosure in scholarly publishing

by Marcel LaFlamme¹ and Natalie Meyers^{1, 2, 3}

1. Association of Research Libraries, Washington, DC
2. Coalition for Networked Information, Washington, DC
3. San Diego Supercomputer Center, University of California San Diego, San Diego, CA

Correspondence: marcel@arl.org

In his essay “Facsimile Machines,” literary scholar Matthew Kirschenbaum (2024) conjures a near-future scenario of scholarly authorship: you begin to write and at once the word processing software you are using, its baked-in AI model drawing on sources you have pointed it toward, chimes in with suggested sentences and paragraphs. As you edit what’s generated or insert a thought of your own, the surrounding text adjusts itself to this new input until, twenty minutes later, your first draft is complete. Yet Kirschenbaum is too astute a historian of labor and technology to stop here, leaving you free to fish away the afternoon. Expectations for scholarly productivity have, in this imagined future, been revised upward such that institutions now look for at least two hundred publications per year. So back to the blinking cursor you go, co-creating another text that other humans will only ever encounter as a secondhand summary.

What anchors the dystopian aspects of Kirschenbaum’s scenario is an almost grotesque surfeit of writing, undertaken to be counted and aggregated rather than read by anyone in particular. To this “tsunami of text” (Kirschenbaum 2023), we want to counterpose another proliferation of language arising from the adoption of AI tools by scholarly authors. Here, we are referring to the production and circulation of AI disclosure statements, a novel feature of scholarly publications that has gained remarkable currency over the past three years. Whether found in a journal article’s methods section, acknowledgments, a supporting information file, or a dedicated metadata field, disclosure statements represent scholarly publishing’s initial attempt to square the disruptive impact of AI with existing norms of authorship and research integrity. By requiring these statements of submitting authors, publishers aim to ensure that the intervention of AI tools does not go unremarked and that editors, reviewers, and readers have the knowledge they need to critically evaluate the works they encounter.

This commentary considers the evolution and limitations of AI disclosure by scholarly authors, as well as reflecting on the implications of this contemporary “imperative to tell” (Manderson et al. 2015, 185).

The public release of ChatGPT in November 2022 touched off a range of responses across the scholarly publishing landscape. A few publishers, like the *Science* family of journals, initially moved to prohibit the use of AI-generated text or images. But a series of guidelines from industry groups (e.g., STM 2023; WAME 2023) soon captured a growing consensus: while AI was neither to be credited as an author nor cited as a source in its own right, the use of AI tools by human authors was to be permitted as long as it was disclosed. Publisher-specific policies on what and where to disclose would follow. Then, with AI use increasingly taking place across multiple stages of the research and writing process, new forms of transparency were needed to provide more granular information on how these tools were being used. Today, generic free-text disclosures like “assisted by AI” are no longer seen as sufficient. There is instead a move toward machine-readable, taxonomically controlled solutions that can be systematized across publisher platforms and mined at scale for metascientific insight.

Emerging approaches to AI disclosure are informed by practices of software citation, use of contributor role ontologies, and expectations around replicability and reproducibility. For instance, the Artificial Intelligence Disclosure Framework (Weaver 2024) structures disclosure statements around the research stages in which an AI tool was used. Authors can select from a list of thirteen possible uses, which notably includes the editing or translation of text but not its drafting. In a similar vein, researchers at IBM have developed the AI Attribution Toolkit (<https://aiattribution.github.io>; see also He, Houde, and Weisz 2025), a reporting standard with a click-through interface that helps authors write an even more precise statement explaining how they used AI, from no use of AI to complex use across research stages and levels of contribution.

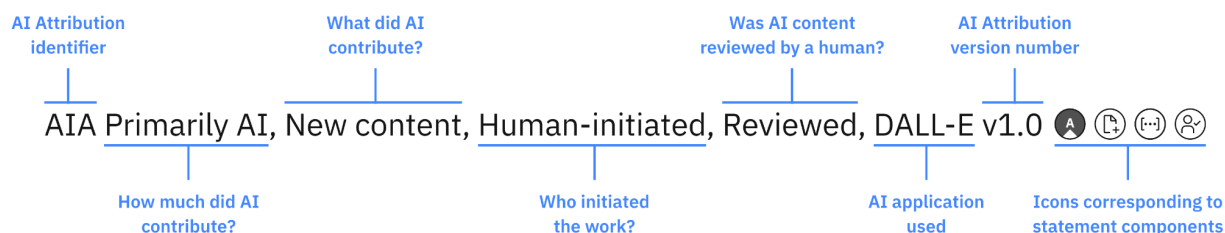


Figure 1. Sample AI attribution statement. (Courtesy of IBM Research)

At the same time, there is a countercurrent that can be seen in initiatives like the AI Free certification (<https://www.aifreecert.com>) and the Authors Guild's Human Authored certification (<https://authorsguild.org/human-authored/>), which offer creators of nonscholarly works a mark of distinction in a contentscape where AI is assumed to be more and more prevalent. Human Authored, which is presently only available to Guild members, allows writers to warrant that the text of a book they have written was not generated by AI. The backstop for this certification is less the technical link from badge to registration number to database than the writer's social connection to the Guild and the representations they have made to it.

What indications do we have that AI disclosures, however they are structured, are achieving their intended purpose? One limitation of disclosure statements is that they rely on authors to understand the scope of relevant activities and to be both truthful and thorough in reporting them. There is some evidence to suggest that scholarly authors are not uniformly complying with publisher policies on AI disclosure. As of this writing, the Academ-AI database (<https://www.academ-ai.info>) contains more than 600 published articles that include language characteristic of generative AI use but do not disclose such use; its creator notes that many more, slightly better edited articles may exist and that detection is likely to become more difficult (Glynn 2024). A second issue relates to the sensemaking that readers engage in when they encounter a disclosure statement. While to our knowledge empirical research has not yet been conducted on reader responses to AI disclosures, an analogous body of research does exist on responses to conflict of interest disclosures in scholarly publications. An analysis of this work (Barrios et al. 2025) found a significant reduction in a publication's perceived credibility when conflicts were disclosed. It is possible that publications disclosing AI use could be subject to similar discounting, which might in turn be expected to affect authors' willingness to disclose (Fang and Lee 2025).

There have been calls for scholarly publishers to move beyond the logic of disclosure in managing authors' use of AI. Avi Staiman (2025) has argued for a risk-based framework, which would involve assigning risk levels to specific tools based on their core uses and maintaining a list of approved tools. Yet, in addition to the practical challenge of monitoring a rapidly changing landscape, the generalist large language models like ChatGPT, Claude, Gemini, and Llama favored by many researchers are difficult to assign a risk level to precisely because of their multifunctionality. Our own view is that disclosure statements can be complemented by the sharing of actual author interactions with AI tools, such as transcripts of prompt

output, as static files in a trusted repository and/or resolvable URLs. This practice, like the sharing of data, code, and protocols already expected by many scientific publications, can also be linked to a trend toward sharing processes more openly in humanities and social science fields like anthropology (Elfenbein, Hoffman, and LaFlamme 2025). Normalizing the sharing of AI interactions will not solve the problem of nondisclosure, but it would permit a richer, more qualitative assessment of the contributions to scholarship made by disclosed AI use—as well as enabling the creative reuse of shared interactions in teaching or further research.

As standards for AI disclosure continue to evolve, there is also a growing awareness that not all uses of AI may need to be disclosed. Voices that previously argued for universal disclosure have recently called for mandatory disclosure to be reserved for uses that are “intentional and substantial” (Resnik and Hosseini 2025, 5), while uses that do not affect the content of research could be considered optional or even unnecessary to disclose. Proponents of this position generally concede that different disciplines may parse these categories differently, but stop short of acknowledging that distinctions between content and form are always local and power-laden. Given the centrality of writerly craft to the production of authority and reputation in social/cultural anthropology (Boyer 2016), textual interventions by AI tools that are considered unnecessary to disclose in other fields may well be seen as vital to disclose by anthropologists. Publishers of anthropological scholarship would therefore do well to set AI disclosure policies with this epistemic diversity in mind, even as anthropologists inside and outside of the academy debate what these disclosures—and the activities they index—mean and ought to mean.

References

Barrios, John M., Filippo Lancieri, Joshua Levy, Shashank Singh, Tommaso Valletti, et al. 2025. “The Conflict-of-Interest Discount in the Marketplace of Ideas.” Working paper, National Bureau of Economic Research. <https://doi.org/10.3386/w33645>

Boyer, Dominic. 2016. “The Necessity of Being a Writer in Anthropology Today.” In *The Anthropologist as Writer: Genres and Contexts in the Twenty-First Century*, edited by Helen Wulff, 21–32. New York: Berghahn.

Elfenbein, Timothy W., Andrew S. Hoffman, and LaFlamme, Marcel. 2025. “Emerging Forms of Open Research in Social/Cultural Anthropology.” Manuscript under review.

Fang, Jingchao, and Mina Lee. 2025. "What Shapes Writers' Decisions to Disclose AI Use?" arXiv, May 27. <https://doi.org/10.48550/arXiv.2505.20727>

Glynn, Alex. 2024. "Suspected Undeclared Use of Artificial Intelligence in the Academic Literature: An Analysis of the Academ-AI Dataset." arXiv, November 20. <https://doi.org/10.48550/arXiv.2411.15218>

He, Jessica, Stephanie Houde, and Justin D. Weisz. 2025. "Which Contributions Deserve Credit? Perceptions of Attribution in Human–AI Co-creation." *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 540. <https://doi.org/10.1145/3706598.3713522>

International Association of Scientific, Technical & Medical Publishers (STM). 2023. "Generative AI in Scholarly Communications: Ethical and Practical Guidelines for the Use of Generative AI in the Publication Process." <https://stm-assoc.org/document/stm-generative-ai-paper-2023>

Kirschenbaum, Matthew. 2023. "Prepare for the Textpocalypse." *Atlantic*, March 8. <https://www.theatlantic.com/technology/archive/2023/03/ai-chatgpt-writing-language-models/673318>

Kirschenbaum, Matthew. 2024. "Facsimile Machines." *Poetics Today* 45(2): 325–29. <https://doi.org/10.1215/03335372-11092977>

Manderson, Lenore, Mark Davis, Chip Colwell, and Tanja Ahlin. 2015. "On Secrecy, Disclosure, the Public, and the Private in Anthropology." *Current Anthropology* 56(S12): S183–190. <https://doi.org/10.1086/683302>

Resnik, David B., and Mohammad Hosseini. 2025. "Disclosing Artificial Intelligence Use in Scientific Research and Publication: When Should Disclosure Be Mandatory, Optional, or Unnecessary?" *Accountability in Research*, March 24. <https://doi.org/10.1080/08989621.2025.2481949>

Staiman, Avi. 2025. "When Declarations Just Don't Cut It: Building a Risk-Based Framework for AI Guidelines in Publishing." *Science Editor* 48(1). <https://doi.org/10.36591/SE-4801-05>

Weaver, Kari D. 2024. "The Artificial Intelligence Disclosure (AID) Framework: An Introduction." *College & Research Libraries News* 85(10), 407.
<https://doi.org/10.5860/crln.85.10.407>

World Association of Medical Editors (WAME). 2023. "Chatbots, Generative AI, and Scholarly Manuscripts: WAME Recommendations on Chatbots and Generative Artificial Intelligence in Relation to Scholarly Publications." <https://wame.org/page3.php?id=106>