

August – 2026

Determining Accountability in AI-Assisted Scholarship: A Report and Recommendations

Jim O'Driscoll¹ and Rory McGreal²

¹Independent Scholar; ²UNESCO Chair in OER, Athabasca University, Canada

Abstract

The proliferation of generative artificial intelligence (GenAI) presents a significant challenge to traditional notions of accountability in scholarly publishing. This report addresses the pressing question of who is responsible for the content of academic work when its creation involves AI assistance. It does so by introducing and extending Erving Goffman's (1981a, b) concept of *production format*, a framework that deconstructs the singular speaker into three distinct roles: *animator*, *author* (herein re-labelled *designer*), and *principal*. By applying this framework to written, scholarly communication, this report analyzes two key problems: the accountability of publishers and platforms as relaying animators and the apparent displacement of the authorial role by GenAI. The analysis draws on recent empirical evidence of journal policy adoption and theoretical applications of Goffman's work to the AI context. It pays particular attention to the implications for open access publishing, where the democratization of knowledge creation must be balanced with the imperative for rigorous accountability. The report concludes that while GenAI can assume a large proportion of the designer function of shaping textual form, the ultimate accountability for scholarly work must remain with a human principal. Detailed recommendations are provided for scholars, publishers (with specific focus on open access journals), and academic institutions to navigate the new landscape of AI-assisted authorship, ensuring transparency and preserving the integrity of academic discourse.

Keywords: generative AI, GenAI, accountability, authorship, scholarly publishing, Goffman, production format, open access

Introduction

The advent of generative artificial intelligence (GenAI) has precipitated a crisis of accountability across numerous domains, none more so than in scholarly publishing. The fundamental academic question, Who said that? has become profoundly complex. For reasons of credit, copyright, legal culpability, and the very pursuit of verifiable knowledge, it is crucial to know who is responsible for a claim and whom to approach for clarification or evidence. Traditionally, the named author(s) beneath a title served this function, signifying the sole linguistic agent responsible for the text's creation and content. GenAI, however, has fundamentally disrupted this long-standing assumption. This report explores this disruption by applying and extending the concept of *production format*, first advanced by the sociologist Erving Goffman, to offer a framework for understanding and navigating accountability in AI-assisted academic work, with particular attention to the context of open access scholarly publishing.

Production Format: A Deconstructed Framework for Authorship

Nearly half a century ago, Goffman (1981a, pp. 144–155; 1981b, pp. 226–229) wrote two papers in which he deconstructed the unitary notion of the speaker, a concept long taken for granted both in communication studies (e.g. Shannon & Weaver, 1949) and in the burgeoning field of speech act theory (e.g. Searle, 1969). In these, he proposed a production format comprising three analytically distinct roles involved in the production of any utterance:

- *Animator*: The individual who physically produces the utterance through the activation of vocal equipment.
- *Author*: The agent who composes the words and selects the syntactic and semantic structure of the utterance.
- *Principal*: The party whose position, beliefs, or stance is represented by the utterance and who is held accountable for it.

In informal conversation, these three roles typically reside in a single person. However, Goffman points out that in other kinds of language production, they are frequently separated. Examples come easily to mind. A politician (principal) delivering a speech may employ a speechwriter (author), while the act of speaking (animation) is their own; a lawyer reading aloud a statement to the press animates words they may have authored, but the position is that of their client (principal).

Goffman's framework was devised for spoken interaction. However, it is readily applicable to written communication. The same concerns of production and reception pertain, with initial animation achieved through the use of a pen or keyboard. To fully address the complexities of scholarly publishing in the age of GenAI, however, the framework requires one key extension and one terminological adjustment.

The extension (only hinted at by Goffman) is to allow for each role to be shared by multiple agents. His authorial role, for instance, is often a collaborative process between a writer and an editor, or, as this report will explore, between a human and an AI platform. The animator role in written communication is almost always shared, requiring the services of a courier—a postal service, a publisher, or an Internet service

provider to make the text accessible to its audience. Even the principal role can be diffused, as when a scholar presents and therefore takes responsibility for another theorist's ideas, or when a politician's utterances are understood to represent a political party.

The terminological adjustment concerns Goffman's author. In common parlance – and indeed the rest of this report – this word denotes not only his author but also his principal. This conflation creates confusion when his framework is applied to the written mode, especially when applied to GenAI, which can perform a compositional role. This report therefore proposes replacing Goffman's author with the term *designer*. This term more precisely captures the function of crafting the textual artifact, determining how to say something without implying the ownership of the underlying ideas, which is the province of the principal. The principal has something to say (meanings), the designer works out how to say it (form), and the animator makes the chosen words manifest. This distinction is crucial as Varoufakis (2023) highlights in his discussion of the ancient Greek principle of *ισεγορία* (isegoria), which not only upholds the right to have one's views considered on their merits regardless of who you are but also separates the substance of an argument from its rhetorical flourish or mode of expression (a task that can be legitimately shared with AI).

A further strand of Goffman's work is relevant here. Recent scholarship has begun to apply his dramaturgical sociology (1959) to the AI context in ways that support and enrich this extended framework. Klowait and Erofeeva (2025) argue that AI in the workplace should be viewed as supplementary rather than substitutive automation. Drawing on Goffman, they reframe AI as a “prop within an ongoing social performance—one whose scripts, stages, and audiences remain irreducibly human.” In their analysis, large language models (LLM) like ChatGPT quietly take over the invisible, routinized tasks that underpin cognitive occupations, namely editing, summarizing, and first-draft production, while leaving humans to enact the highly visible or relational facets that sustain professional prestige. This perspective aligns entirely with the proposed distinction between the AI as a designer (handling routinized textual formulation) and the human as the principal (enacting professional competence and ultimate accountability). The human's role shifts from being the sole designer to being the manager and approver of the design process, a role that requires a new form of AI literacy to ensure the final output aligns with their intellectual stance.

Problem 1: The Animator as Additional Principal in the Open Access Context

A central problem for scholarly publishing concerns the accountability of what we may term relaying animators, namely the publishers, platforms, and Internet service providers that facilitate the dissemination of a text. While these entities often position themselves as neutral messengers, their editorial power and gatekeeping function suggest they also partake of the principal role. This issue is particularly acute for open access (OA) journals, which often operate with fewer resources and may be more vulnerable to both the risks and rewards of AI integration.

A 2025 incident involving the BBC and historian Rutger Bregman's Reith Lectures illustrates this quandary. In one of these commissioned series of lectures, delivered live to an audience, Bregman described a political figure as “the most openly corrupt president in American history.” But when the recording of this lecture was broadcast, this phrase was removed by the BBC on legal advice (Yahoo, 2025). In the live context, Bregman was clearly the sole principal. In the broadcast version, however, the BBC's editorial intervention

made it a shared principal, accountable for the final content its audiences received. This raises a critical question: to what extent does editorial control confer accountability?

The editorial power of publishers, from major academic houses to niche journals, makes them de facto additional principals. For example, an editor's discomfort with an author's opinion can effect real changes in the content of an article. This reflects a sense of ownership over the publication's entire content. This is not merely a theoretical concern. Legal history, such as the mid-20th century obscenity trial of *Lady Chatterley's Lover* being brought not against its author but rather its publisher Penguin Books, demonstrated that society often holds these relaying animators accountable. In the scholarly context, the reputation of a journal or press is inextricably linked to the works it publishes.

However, granting relaying animators full principal status is dangerous. If powerful publishers and platforms are held legally accountable for every idea they disseminate, a chilling effect on academic freedom and diversity of thought becomes inevitable. This risk is amplified in an era of LLMs, which are trained on existing corpora and tend to reinforce dominant ideologies. The solution may be to recognize them as additional principals, with a tiered accountability: the primary accountability for the truthfulness and originality of claims rests with the human author, while the publisher is accountable for the procedural integrity of the publication—the rigour of peer review, the transparency of editorial processes, and the adherence to ethical guidelines. This distinction is vital for OA journals, which are committed to disseminating knowledge broadly and must safeguard against any perception that open access equates to lower standards.

Recent empirical research has underscored the urgent need for publishers to formalize their role as additional principals through clear policies. In a cross-sectional review of vascular surgery journals, Shen et al. (2025) found that while 72% had explicit GenAI policies, significant inconsistencies remained. All journals prohibited AI authorship and affirmed human accountability, a non-negotiable baseline. However, only 56% specified where disclosure of AI use should be placed, and only 52% addressed AI-generated images. Crucially, higher-impact journals (Q1/Q2 in SCImago Journal Rank) demonstrated more comprehensive policies than lower-tier journals. This disparity is a major concern for the OA landscape, which includes many new, lower-profile and scholar-led journals that may lack the editorial infrastructure to develop robust policies. Without clear guidance from these additional principals, the integrity of the scholarship they publish, and by extension the reputation of the OA movement itself, is put at risk.

AOSIS (African Online Scientific Information Systems) has provided a model response. In late 2025, it made the inclusion of AI policies compulsory for all journals in its indexing service (SciELO SA), reminding authors to disclose their use of AI-assisted technology to ensure transparency, accuracy, and trust. Furthermore, AOSIS announced its transition to the CRediT (Contributor Roles Taxonomy) authorship statement, a move that dramatically increased transparency by specifying each author's contribution to a work. This taxonomy, which includes categories like 'conceptualization,' 'methodology,' and 'writing—original draft,' operationalizes the deconstruction of the author role. It provides a practical tool for distinguishing between the human principal's contributions and tasks that may have been assisted by AI, thereby reinforcing accountability.

Problem 2: The Author (Designer) as Principal in the Age of GenAI

For centuries, copyright law and academic convention have conflated the roles of designer and principal under the single term author. This conflation is now becoming untenable. With appropriate prompting, LLM can generate full texts, effectively assuming the designer role. This is a core disruption of GenAI. The question then becomes: Who is the principal? If the author in the conventional sense no longer composes the text, can they still be held accountable for its claims?

The use of AI in this context has sparked widespread misgivings about the death of creativity and thought in academic work. However, as Gunkel (2025) argues, this perspective may be misguided, the development having one notably positive effect. The absence of a singular, intentional authorial voice can liberate readers from the hermeneutic task of excavating a text for the author's 'true meaning', encouraging instead a focus on the text's effects and the reader's own interpretation. This recipient-centered approach aligns with contemporary linguistic pragmatics, which has become increasingly suspicious of the notion of speaker/writer intention (e.g., Terkourafi, 2021, pp. 41–43). It also democratizes critique: chiming well with Varoufakis' account of isegoria, a text can only be judged on its merits, not on the authority of its creator.

Nevertheless, the need for an identifiable principal remains paramount, especially in the academic domain. Unlike folklore or popular songs, where accountability is largely irrelevant, academic work directly informs further research, policy, and real-world actions. A principal is essential for follow-up, clarification, and, most critically, for responsibility if claims are found to be flawed or fraudulent. The case of the retracted measles, mumps, and rubella (MMR) vaccine study, which could have caused real-world harm, has underscored the danger of research without a clear human party to answer for it. And while a student's use of GenAI for writing assignments complicates assessment, to deprive them of available tools would be perverse. The solution is not prohibition, but a reconfiguration of what it means to be an author in the traditional sense.

This reconfiguration demands that the human user of GenAI, by deciding which AI-generated content to incorporate, edit, and ultimately stand behind, retains the principal role. The AI serves as a sophisticated advisor or mentor, whose contributions can be acknowledged but who cannot be held accountable for the final published work. Mank's (2025) editorial in *Evolution Letters* powerfully reinforces this point. It identifies accountability as the first of three principles (alongside voice and trust) that must be preserved. Mank is unequivocal that "ultimately, the authors of a paper are accountable for the data, analysis, and manuscript itself. . . . The human writing the prompts and doing the analysis is ultimately responsible for checking that anything they implement from LLMs is appropriate and correct". She even models this transparency by disclosing her own use of ChatGPT to generate a section of the editorial, demonstrating that AI use can be acknowledged without diminishing authorial accountability.

Mank (2025) also introduces the critical concept of voice, arguing that we cannot allow LLMs to override our own voice in manuscripts and reviews. This is the practical manifestation of the principal-designer distinction. The AI (designer) may suggest a turn of phrase, but the human (principal) must engage critically with that output, revising it to align with their own argument, priorities, and authentic scholarly identity. This process ensures that the final text is not merely a competent piece of writing, but a genuine representation of the principal's intellectual stance. And, in an era where everyone may have access to the

same AI tools, the ability to cultivate and assert a unique scholarly voice may become more, not less, important. For this reason, it is important that the human writer not only retains the principal role but also shares (with AI) the designer role.

The Special Imperative for Open Access Journals

The framework of production format has distinct and profound implications for open access publishing. The OA movement is founded on the principles of accessibility, equity, and the democratization of knowledge. The integration of GenAI, viewed through this framework, presents both a powerful opportunity and a significant threat to these principles.

There is an opportunity for enhancing democratic participation when the AI acts as a designer. This can lower barriers for scholars who are not fluent in written English or who are from under-resourced institutions. As noted by Mank (2025), LLMs are “remarkably effective at helping reduce the cognitive burden of writing in a language with which they are less comfortable”. By assisting with the formulation of ideas into publishable prose, AI can help a more diverse range of scholars assume the role of principal, contributing their unique perspectives to the global academic conversation. This affordance aligns perfectly with the OA goal of amplifying marginalized voices. Furthermore, some envision AI transforming journals from metrics-driven repositories into vibrant knowledge communities that facilitate meaningful discourse and collaborative learning, further enhancing the accessibility and utility of OA content.

However, there is a real possibility that GenAI can be used to undermine trust and equity. The same lack of resources that makes OA journals attractive to scholars from diverse backgrounds can also make them more vulnerable to the risks of GenAI. The above-cited study of vascular surgery journals found that lower-impact journals (which often overlap with smaller, newer OA journals) had significantly less comprehensive AI policies. This policy gap creates a dangerous vulnerability. Such journals may become targets for ‘paper mills’ that use AI to generate large volumes of low-quality, nonsensical, or even plagiarized content—so-called AI slop. This influx threatens to overwhelm editorial systems, devalue a journal’s brand, and erode the trust of readers, who are the ultimate beneficiaries of OA. If open access journals become perceived as a haven for low-quality, AI-generated work, the entire OA movement suffers a reputational blow. The trust that is the foundation of scholarly communication is shattered when readers cannot be sure if a human principal stands behind the claims being made.

Therefore, for OA journals, adopting and enforcing clear GenAI policies is not merely an ethical nicety; it is an existential imperative. They must embrace their role as additional principals with vigour. This means developing policies that are as robust as those of top-tier journals, even with limited resources. It means providing clear, simple guidelines for authors on disclosure, perhaps by adopting a version of the CRediT taxonomy to broadly separate human and AI contributions, without imposing unnecessary tracking that could make the process more onerous than is necessary. It also means investing in editor and reviewer training to spot the hallmarks of uncritical AI use. By doing so, OA journals can protect their integrity and demonstrate that accessibility and rigorous accountability are not mutually exclusive but are instead mutually reinforcing pillars of a healthy scholarly ecosystem.

Conclusions and Recommendations

Goffman's (1981a, b) production format, extended and adapted here, provides a powerful lens through which to view the challenges GenAI poses to academic accountability. Recent theoretical work has confirmed the value of Goffman's approach, while empirical studies have highlighted the urgent but inconsistent response from publishers. A model for an AI-assisted scholarly work can be visualized as follows.

- **Principal role:** The named human author(s), who determine(s) the core meanings, arguments, and findings to be communicated, and who ultimately stand(s) by the work they disseminate.
- **Designer role:** A collaboration between the human author(s) and GenAI, working together to determine the optimal form, structure, and wording for the argument. The human maintains their voice and critically evaluates all output.
- **Animator role:** A chain including (a) the human author(s), typing and issuing the text; (b) the GenAI interface, generating text on screen; (c) the Internet service provider, transmitting the text; and (d) the publisher, formatting and disseminating the final work.

This model is necessarily a simplification; any writer knows that the process of writing down what you think you want to say helps find out what it is that you really want to say. The process of writing is recursive, with design and animation feeding back into the principal's meaning-making. However, it establishes a clear locus of accountability, namely the human principal. Based on this analysis, the following recommendations are made for different stakeholders in the scholarly publishing ecosystem.

Recommendations

For Scholars and Authors

- **Disclose use.** When using GenAI in the design of a scholarly work, its use must be transparently disclosed. A statement in the acknowledgements or a dedicated methods section is appropriate.
- **Assume principalship and guard your voice.** Authors must understand that using GenAI does not transfer accountability. By submitting a work for publication, they affirm their role as the principal, taking full responsibility for its content, including any text or references generated with AI assistance. They must critically engage with AI output to ensure it reflects their own scholarly voice and does not introduce errors or biases.
- **Distinguish idea from form.** Be aware of the distinction between the ideas (principal) and their formulation (designer). The ultimate intellectual contribution, and the responsibility for it, rests with the human author.

For Publishers and Journals (Including Open Access)

- Establish and enforce clear, standardized policies. Develop and enforce clear policies on the use of GenAI. These policies must explicitly state that AI cannot be listed as an author (principal) and must require full disclosure of its use.
- Re-define editorial accountability. Openly acknowledge the role of the publisher as an additional principal, accountable for the overall quality and integrity of what they publish. This accountability includes ensuring that submissions meet ethical and scholarly standards through a robust peer-review process, regardless of whether they are AI-assisted.
- For open access journals specifically, recognize that robust AI policies are essential for maintaining trust and competitiveness. Proactive policy development is a form of quality assurance that protects the OA brand.
- Update author guidelines. Provide clear, practical guidance to authors on how to disclose GenAI use.

For Academic Institutions and Educators

- Integrate AI literacy into curricula. Teaching students how to use GenAI responsibly and critically is essential. This includes helping them understand their role as principals when they submit AI-assisted work for assessment and ensuring they can critically evaluate AI output.
- Develop institutional guidelines. Formulate guidelines that clarify the responsible use of GenAI in research and teaching, emphasizing the enduring importance of human accountability and intellectual integrity.
- Foster a culture of transparency. Encourage open discussion about the use of these tools to demystify them and to develop best practices collaboratively, ensuring that the focus remains on the pursuit of knowledge rather than the policing of tools.

By embracing a more nuanced understanding of authorship, one that separates the creation of form from the ownership and accountability for ideas, the academic community can harness the power of GenAI while safeguarding the foundational principles of scholarly integrity.

Acknowledgements

After using Generative AI (DeepSeek V-3.2), we reviewed and edited the content as necessary and take full responsibility for it.

References

- Goffman, E. (1959). *The presentation of self in everyday life*. Anchor Books.
- Goffman, E. (1981a). Footing. In *Forms of talk* (pp. 124–159). University of Pennsylvania Press. (Original work published 1979) <https://www.scribd.com/document/535335051/Goffman-1979-Footing-in-Forms-of-Talk-1981>
- Goffman, E. (1981b). Radio talk. In *Forms of talk* (pp. 197–330). University of Pennsylvania Press. <https://archive.org/search?query=external-identifier%3A%22urn%3A%3Aformsoftalkoogoff%3A%3Acpdf%3A6ad3d5a7-b440-42df-a50f-50d96ea200fb%22>
- Gunkel, D. J. (2025). AI personhood. In *A companion to applied philosophy of AI*. Wiley. <https://doi.org/10.1002/9781394238651.ch27>
- Klowait, N., & Erofeeva, M. (2025). The presentation of self in the age of ChatGPT. *Frontiers in Sociology*, 10, 1–13. <https://doi.org/10.3389/fsoc.2025.1614473>
- Mank, J.E. (2025). Editorial. Accountability, voice, and trust: Responsible use of GenAI in scientific publishing. *Evolution Letters*, 9(4), 381–382. <https://doi.org/10.1093/evlett/qrafo27>
- Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge University Press. https://books.google.ca/books/about/Speech_Acts.html?id=t3_WhfknvFOC&redir_esc=y
- Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. University of Illinois Press. https://pure.mpg.de/rest/items/item_2383164_3/component/file_2383163/content
- Shen, F., Almeida, V. F. A., Ha, J. Y., Ozmen, B. B., & Duraes, E. F. R. (2025). Prevalence and characteristics of generative AI policies in vascular surgery journals: A cross-sectional review. (2025). *Annals of Vascular Surgery* 123: 119-129. <https://doi.org/10.1016/j.jvs.2025.06.012>
- Terkourafi, M. (2021). Inference and implicature. In *The Cambridge handbook of sociopragmatics* (pp. 30–47). Cambridge University Press. <https://doi.org/10.1017/9781108954105.004>
- Varoufakis, Y. (2023). *Technofeudalism: What killed capitalism*. Penguin Random House. <https://www.penguinrandomhouse.ca/books/751443/technofeudalism-by-yanis-varoufakis/9781685891237>
- Yahoo (2025). Reith lecturer accuses BBC of censoring his speech by removing line about Trump. *Yahoo News*. https://uk.news.yahoo.com/reith-lecturer-accuses-bbc-censoring-141029037.html?guccounter=1&guce_referrer=aHRocHM6Ly93d3cuYmluZy5jb2ov&guce_referrer_sig=AQAAAF

