



Generative AI and Scholarly Publishing

by Britt Amell | 14 November 2024 | English, Insights and Signals Reports



Lisez-le en français.

This insights and signals report was written by Brittany Amell, with thanks to John Willinsky, John Maxwell, and William Bowen for their feedback and contributions.

At a Glance

| | |
|-------------------------------|------------------------------------------------------------------------------------------------------------------------------|
| Insights & Signals Topic Area | Generative AI and Scholarly Publishing |
| Key Participants | Ithaka S+R, cOAlition S, DOAJ, European Commission Directorate-General for Research and Innovation, Public Knowledge Project |
| Timeframe | 2022 – Present |
| Keywords or Key Themes | Generative AI, open scholarship, trust, credibility, open access |

Summary

This insights and signals report continues OSPO’s review of the evolving dialogue on the implications generative AI has for open scholarship / open access publishing.

“Generative AI” refers to a class of algorithms that guide the creation of various types of content ([Dobrin, 2023](#)). Sometimes referred to in the same breath as [ChatGPT](#), [DALL·E 2](#), or [Gemini](#), we are already starting to see the impact of generative AI on scholarly publishing, despite its relatively recent introduction.

With growing interest in the implications generative AI has for open access and scholarly communication more broadly, this report centers around recent discussions of the potential risks and opportunities for the sector. *Interested in other Insights and Signals Reports focused on AI? You can find them [here](#) and [here](#).*

Items discussed in this report include:

- A [recent announcement](#) from Ithaka S+R regarding a new research project focused on generative AI and scholarly publishing
- A [guest post](#) by Shen and Ball (DOAJ) on the Scholarly Kitchen regarding a surge in retracted articles and the crisis of trust for open access journals amidst innovations in generative AI
- A [proposal](#) put forward by cOAlition S regarding responsible publishing
- The potential applications of the FAIR principles to policy development in response to generative AI, as well as the overlap between these principles and those referenced in a [recent guide to responsible use of AI](#) in research

released by the European Commission Directorate-General for Research and Innovation

- The **Publication Facts Label** (PFL—also: [here](#)), as well as a **recent announcement from the Public Knowledge Project** regarding their trial of the PFL for journals operating on OJS (v 3.3 or higher)

Scholarly Publishing Research Group Announces New Project Exploring Implications of generative AI

Research group **Ithaka S+R** recently announced its **intention to undertake a new research project focused on generative AI and scholarly publishing**: “Rapidly changing user needs and expectations, the potential of generative AI to mitigate stubborn systemic challenges in the scholarly publishing industry, and an awareness of the risks generative AI poses to expert knowledge demand that we find time for deep reflection about what generative AI *means* for scholarly publishing as a sector and its value as a component of the shared infrastructure that supports scholarly and scientific research and communications” (Ruediger and Bergstrom, 2024). This project will consider the opportunities, risks, and strategic implications generative AI could have for scholarly publishing. It builds on Ithaka S+R’s 2023 study, “**The Second Digital Transformation of Scholarly Publishing**,” which examined shared needs for scholarly communication infrastructures in light of digital transformations (Bergstrom, Rieger, and Schonfeld, 2024).

As Shen and Ball (2024) **write in their guest post for The Scholarly Kitchen**, while threats to trust and credibility have always been top-of-mind for organizations like the Directory of Open Access Journals (DOAJ), these threats have taken on an added urgency. Noting a surge in retracted articles that hit a record-breaking number of 10,000 in 2023, Shen and Ball (2024) share how there has been an additional need to take further steps to safeguard trust and credibility in an era of generative AI.

For the DOAJ, one such step has included the formation of a team dedicated to investigating “suspected instances of questionable practices.” These questionable practices are flagged either by members of the DOAJ community more broadly, or by those participating in the evaluation process of a journal’s application for inclusion in the DOAJ. Once flagged, team members closely review a journal’s published articles,

as well as the composition and competence of its editorial board and its peer review practices alongside several other factors.

“As predatory practices continue to evolve, our investigations are becoming increasingly complex,” write Shen and Ball (2024). “We sometimes consult external subject matter experts for their advice. In 2023 alone, we carried out a total of 409 investigations into journals and publishers, many of which resulted in exclusions from DOAJ of at least one year.”

Supporting publishers and the public in assessing the credibility of a publication is something those at the Public Knowledge Project (PKP) have also thought a lot about. For instance, the Publication Facts Label originally conceptualised by John Willinsky (INKE partner and founder of the Public Knowledge Project; discussed in more detail both [here](#) and [here](#)) is one such example. Based on the Nutrition Facts Label—that well-known chart found on food packaging in both Canada and the United States—the Publication Facts Label (or PFL for short) consolidates eight standards into an at-a-glance guide that publishing platforms can use to present the integrity of a publication to a wide audience (Willinsky and Pimental). PKP [recently announced](#) the trial of the Publication Facts Label for journals using the Open Journal Systems (v 3.3 or higher). By installing a plug-in, as explained [here](#), the PFL can be displayed on an article’s landing page automatically.

cOAlition S Wraps up Responsible Publishing Proposal

As Ithaka S+R prepares to move ahead with its project, [cOAlition S](#) (associated with [Plan S](#)) has wrapped up the consultation phase for their proposal entitled “[Towards responsible publishing](#).”

“Driven by the same ‘duty of care for the good functioning of the science system’ that inspired Plan S, the funders forming cOAlition S are now exploring a new vision for scholarly communication; a vision that holds the promise of being more effective, affordable, and equitable, ultimately benefiting society as a whole,” writes Bodo Stern (Chief of Strategic Initiatives at the Howard Hughes Medical Institute) and Johan Rooryck (Executive Director of cOAlition S) in a [blog post announcing the proposal](#).

cOAlition S revised the proposal based on the feedback they received between November 2023 and April 2024, and shared the revised proposal with a gathering of cOAlition S funders in June.

Feedback on the proposal has been made publicly available by cOAlitionS [here](#).

The “Towards responsible publishing” proposal puts forward a set of principles that can be used to guide decisions regarding how to support “the dissemination of research in a responsible, equitable, and sustainable way” (Stern et al. 2023, 2)—a phrase that evokes the key priorities named in the [European University Association’s Open Science Agenda 2025](#). The Association’s Open Science Agenda, which was released in February 2022 ahead of its upcoming conference and general assembly in 2025-2026, outlines several priorities and aims, including the goal for all of Europe’s universities to be a part of a “just scholarly publishing ecosystem” by 2025 ([Gaillard, 2022](#)).

A just scholarly publishing ecosystem is described by Vinciane Gaillard (Deputy Director for Research and Innovation, European University Association) as one that is “transparent, diverse, economically affordable and sustainable, technically interoperable, and steered by the research community and its institutions through coordinated policies” ([slide 4](#)).

Elsevier launches ‘ScopusAI’

Earlier in 2024, Elsevier announced the release of ScopusAI, a subscription-based AI tool offered to institutions. According to Elsevier’s [website](#), ScopusAI serves as “an expert guide” researchers can use to navigate “the vast expanse of human knowledge in Scopus.” In addition to summarizing literature that is available through Scopus, the tool is apparently also able to “pinpoint” what Elsevier refers to as “white space” in the literature so that researchers can ostensibly better identify the kinds of contributions they can make. Concerningly, the summaries generated by the tool link up to what Elsevier and the ScopusAI tool decide are “foundational” documents for a topic—these are “high-impact papers most commonly cited by the papers used in the summaries.” In addition to algorithmic decisions made regarding which papers are considered foundational, the tool also offers the opportunity to “discover experts” in an area. However, the tool only considers papers and profiles

that are already in Scopus, which is already known as a database that overrepresents scholars and scholarship from Europe, Oceania and North America when compared to scholars and scholarship from other regions of the world regions (Asubiaro et al. 2024).

Key Questions and Considerations

The insights and signals outlined above and [elsewhere](#) indicate that the integration of generative AI in scholarly publishing presents both opportunities and challenges, as well as significant implications for the quality, integrity, and accessibility of scholarly outputs. Several questions for policy development arise, including:

- What place, if any, is there for generative AI in open access publishing?
- How might open access publishers guide the responsible, ethical, and credible use of generative AI?

For instance, John Willinsky (Founder, Public Knowledge Project and INKE partner) has shared with us that PKP is actively exploring the ways in which AI based on large language models might support the sustainability of open access publishing:

The principal focus of this work is to establish if LLMs can be sufficiently tuned to reliably automate HTML and JATS XML markup of author manuscripts (given that such markup currently requires technical skills or payments that exceed the capacity of most Diamond OA journals. This work has reached an initial proof of concept stage, with further work continuing around its comparative value (given other tools) and ways of incorporating and sustaining such a markup service in the editorial workflow. (Read John Willinsky's comment in full below.)

In addition to these questions, readers might also consider what, if any, existing key lessons and insights from the OA movement, discourse, research, and literature might be applied to the evolving landscape of generative AI. For instance, one example might be the FAIR principles for data management and stewardship ([Wilkinson et al. 2016](#))—originally envisioned as a way to support the reuse of scholarly data by ensuring it is findable, accessible, interoperable, and reusable. The FAIR principles have been re-interpreted to apply to software, workflows, tools, algorithms, and, increasingly, AI models ([Huerta et al. 2023](#)). In a related vein, the

European Commission Directorate-General for Research and Innovation recently **released a set of principles to help guide the responsible use of generative AI in research**. These principles could be said to overlap with the FAIR principles, offering further opportunity for reflection. (For reference, the four key principles are reliability, honesty, respect, and accountability.)

Comments from the INKE Partnership

Response from John Willinsky (Founder, Public Knowledge Project):

While there are reasons to be concerned about recent advances in AI, academics also have a responsibility to explore the potential contributions and advances that AI may hold for research and scholarship. The **Public Knowledge Project** has for some time looked to AI to solve pressing issues around resource equity and quality in scholarly communications, if only with limited success. It is now engaged in research on the ability of Large Language Models to address the long-standing challenge of developing a sustainable means for Diamond OA journals to publish in the standard formats of HTML and PDF, as well export files in JATS XML. The principal focus of this work is to establish if LLMs can be sufficiently tuned to reliably automate HTML and JATS XML markup of author manuscripts (given that such markup currently requires technical skills or payments that exceed the capacity of most Diamond OA journals. This work has reached an initial proof of concept stage, with further work continuing around its comparative value (given other tools) and ways of incorporating and sustaining such a markup service in the editorial workflow.

References

Ahari, Juni. 2024. "Generative AI and Scholarly Publishing." *Ithaka S+R* (blog). April 23, 2024. <https://sr.ithaka.org/blog/generative-ai-and-scholarly-publishing/>.

Asubiaro, Toluwase, Sodiq Onaolapo, and David Mills. 2024. "Regional Disparities in Web of Science and Scopus Journal Coverage." *Scientometrics* 129 (3): 1469–91. <https://doi.org/10.1007/s11192-024-04948-x>.

Bergstrom, Tracy, Oya Y. Rieger, and Roger C. Schonfeld. 2024. "The Second Digital Transformation of Scholarly Publishing: Strategic Context and Shared Infrastructure." Ithaka S+R. <https://doi.org/10.18665/sr.320210>.

Chiarelli, Andrea, Ellie Cox, Rob Johnson, Ludo Waltman, Wolfgang Kaltenbrunner, André Brasil, Andrea Reyes Elizondo, and Stephen Pinfield. 2024. "'Towards Responsible Publishing': Findings from a Global Stakeholder Consultation." cOAlition S. Zenodo. <https://doi.org/10.5281/zenodo.11243942>.

Directorate-General for Research and Innovation. 2024. "Living Guidelines on the Responsible Use of Generative AI in Research (Version 1)." Brussels: European Commission. https://research-and-innovation.ec.europa.eu/document/download/2b6cf7e5-36ac-41cb-aab5-0d32050143dc_en?filename=ec_rtd_ai-guidelines.pdf.

Dobrin, Sidney I. 2023. "Talking about Generative AI: A Guide for Educators." Version 1. Broadview Press. <https://sites.broadviewpress.com/ai/talking/>.

Gaillard, Vinciane. 2022. "Encouraging/Supporting Sustainability in the Diamond Action Plan Community." Presented at the 2022 Diamond Open Access Conference, September.

Huerta, E. A., Ben Blaiszik, L. Catherine Brinson, Kristofer E. Bouchard, Daniel Diaz, Caterina Doglioni, Javier M. Duarte, et al. 2023. "FAIR for AI: An Interdisciplinary and International Community Building Perspective." *Scientific Data* 10 (1): 487. <https://doi.org/10.1038/s41597-023-02298-6>.

Shen, Cenyu, and Joanna Ball. 2024. "DOAJ's Role in Supporting Trust in Scholarly Journals: Current Challenges and Future Solutions." *The Scholarly Kitchen* (blog). June 6, 2024. <https://scholarlykitchen.sspnet.org/2024/06/06/guest-post-doajs-role-in-supporting-trust-in-scholarly-journals-current-challenges-and-future-solutions/>.


Stern, Bodo, and Johan Rooryck. 2023. "Introducing the 'Towards Responsible Publishing' Proposal from cOAlition S | Plan S." sOApbox: A Plan S Blog (blog).

October 31, 2023. <https://www.coalition-s.org/blog/introducing-the-towards-responsible-publishing-proposal-from-coalition-s/>.


Stern, Bodo, Zoé Ancion, Andreas Björke, Ashley Farley, Marte Qvenild, Katharina Rieck, Jeroen Sondervan, et al. 2023. “Towards Responsible Publishing: Seeking Input from the Research Community to a Draft Proposal from cOAlition S,” October. <https://doi.org/10.5281/ZENODO.8398480>.

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. “The FAIR Guiding Principles for Scientific Data Management and Stewardship.” *Scientific Data* 3 (1): 160018. <https://doi.org/10.1038/sdata.2016.18>.

Search



Archives



Categories

Community News

English

French

Insights and Signals Reports

Observations

Observations and Responses

[Policies](#)

[Responses](#)

[Uncategorized](#)

Tags

[Berlin Declaration / Déclaration de Berlin](#)

[Bethesda Statement / Déclaration de Bethesda](#) [bibliodiversity / bibliodiversité](#)

[Budapest Statement / Déclaration de Budapest](#) [Canada](#) [Canadiana.org](#)

[Canadian government/le gouvernement du Canada](#) [CAPOS](#) [CARL / ABRC](#)

[collaboration](#) [Compute Canada / calcul Canada](#) [copyright / droits d'auteurs](#)

[CRKN / RCDR](#) [data management / gestion des données](#)

[digital scholarship / version numérique](#) [en français / French](#) [english](#)

[English / en anglais](#) [Federation for the HSS / Fédération des sciences humaines](#)

[funding agencies / organismes de financement](#) [genAI](#)

[generative artificial intelligence / l'intelligence artificielle générative](#)

[identity management / gestion de l'identité](#) [implementation / mise en oeuvre](#) [INKE](#)

[international policy / politique internationale](#) [Naylor Report / le rapport Naylor](#)

[open access / libre accès](#) [open data / données ouvertes](#)

[open education / éducation ouverte](#) [open government / gouvernement ouvert](#)

[open infrastructure / infrastructure ouverte](#) [open scien](#)

[open science / science ouverte](#) [open source software / les logiciels libres](#) [ORCID](#)

[peer review / critique des pairs](#) [PKP](#) [Plan S](#) [Plan S update / mise à jour du Plan S](#)

[policy / politique](#) [promotion et titularisation](#) [publishing / édition](#) [RDC / DRC](#)

[RDM](#) [RECODE](#) [recommendations / recommandations](#) [reports / les rapports](#)

[repositories / les dépôts](#) [research evaluation / l'évaluation de la recherche](#)

[research evaluation / évaluation de la recherche](#)

[research libraries / les bibliothèques de recherche](#)

[research output / les résultats de la recherche](#)

[research security / sécurité de la recherche](#) [RPT / révision](#)

[scholarly communication / la communication savante](#) [SFU Library / Bibliothèque](#)

[social media / les médias sociaux](#) [Tri-Agency / des trois organismes](#) [UK](#)

[UK / Royaume-Uni](#) [UNESCO](#) [UVic Libraries](#) [Érudit](#)



This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](#).

