AUTHORS:
Brenda D. Wingfield[1]
Beverley J. Wingfield[2]

AFFILIATIONS:
[1]Department of Biochemistry, Genetics and Microbiology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria, South Africa
[2]Department of Financial Management, University of Pretoria, Pretoria, South Africa

CORRESPONDENCE TO:
Brenda Wingfield

EMAIL:
Brenda.Wingfield@up.ac.za

HOW TO CITE:
Wingfield BD, Wingfield BJ. The hidden cost of open access: Artificial intelligence, paywalls and the risk of knowledge inequity. S Afr J Sci. 2025;121(11/12), Art. #23053. https://doi.org/10.17159/sajs.2025/23053

ARTICLE INCLUDES:
☐ Peer review
☐ Supplementary material

KEYWORDS:
artificial intelligence, open access, journal paywalls, large language models, LLMs, knowledge equity

PUBLISHED:
26 November 2025

The hidden cost of open access: Artificial intelligence, paywalls and the risk of knowledge inequity

Significance:

This Commentary underscores the urgency of rethinking how artificial intelligence systems and publishing models intersect. Rather than focusing solely on efficiency or access, we call attention to structural inequities that shape who is visible in the scholarly record. By highlighting the systemic risks of exclusion, we argue for proactive collaboration between policymakers, publishers and technologists to design frameworks that safeguard equity and inclusivity in the future of academic research.

Introduction

Artificial intelligence (AI) systems[1], such as ChatGPT and other large language models (LLMs), are increasingly utilised in academic research, and have transformed the way scholars and students interact with information. These systems offer quick summarisation, language support and thematic integration, improving access to and understanding of scholarly literature. However, one limitation of AI systems that is often overlooked is their reliance on open-access academic literature, on which they almost exclusively depend. This dependence has significant implications for knowledge equity, economic inequality and the accuracy of scientific discourse.

The most immediate and worrying impact of AI's limited access to open-source content is the invisibility of paywalled literature. Many foundational or influential studies are still published in traditional subscription-based journals, including those managed by major academic publishers such as Elsevier, Springer Nature, Wiley and Taylor & Francis. We recognise that there has been a rapid growth of green open access with associated rights and retention strategies, preprints are becoming more common, and many libraries have negotiated text and data mining clauses. However, many articles are still inaccessible to AI systems that are not legally or technically permitted to train on copyrighted or restricted content. This restriction is worsened by legal and technical barriers that hinder text and data mining of subscription-based content[2], creating a structural gap in AI's ability to access the complete scholarly record.

Consequently, AI-assisted literature reviews, summaries and citations depend heavily on open-access sources, which creates a systematic bias in how knowledge is represented. Key findings published in paywalled journals are often excluded or understated, even when they are central to the scholarly consensus on a topic. As more researchers, students and editors turn to AI tools to support or initiate literature reviews, the exclusion of paywalled research could distort the development of entire research narratives. Traditional institutional subscriptions often leave many researchers –particularly in the Global South – with limited access to scholarly literature, even though the pirate site Sci-Hub provides access to a large portion of this paywalled scholarly literature.[3]

The consequence of this exclusion could be diminished visibility and influence of paywalled publications, effectively reducing their citations and role in shaping academic debates. It also threatens the epistemological integrity of knowledge creation – what is considered 'known' or 'important' increasingly depends on accessibility rather than merit.

Although open access is often portrayed as a democratising force in science, the expense of article processing charges (APCs) remains a significant barrier, especially for researchers in developing countries. APCs can vary from a few hundred to several thousand US dollars and are usually billed in currencies that are either inaccessible or devalued compared to those of the Global South. This creates a structural disadvantage for scholars in Africa, South America, Southeast Asia, and parts of Eastern Europe. Klebel and Ross-Hellauer[4] demonstrated that APC-funded publishing predominantly benefits researchers in wealthier nations, further reinforcing structural inequalities.

Case study: The South African context

South Africa demonstrates how global inequalities in scholarly publishing unfold locally. Library consortia such as SANLiC (South African National Library and Information Consortium) negotiate subscription access with major publishers. Yet, rising costs mean that even well-funded universities cannot guarantee complete disciplinary coverage. Simultaneously, national policies from the National Research Foundation (NRF) and the Academy of Science of South Africa (ASSAf) strongly support open-access dissemination, but the reliance on APCs introduces new barriers.

APCs for leading international journals often surpass ZAR40 000–ZAR60 000 per article, making them unaffordable for many researchers without subsidies. To put this into perspective, the NRF's full-cost bursary for a PhD student is ZAR175 000 annually, while a typical PhD scholarship in the USA is valued at USD25 000 (≈ZAR450 000). Consequently, a single APC can consume a significant portion of a South African student's yearly support, whereas in the Global North, it constitutes a much smaller share.

This dual challenge – restricted subscription access and prohibitive OA publishing costs – undermines visibility, discoverability and participation in global scholarship. Early-career researchers and postgraduate students are especially disadvantaged, with many unable to publish at all. These realities emphasise the urgent need for policies that address both access and affordability, ensuring South African researchers are not systematically sidelined in an era of AI-assisted scholarship.

Even when research grants are secured, these APCs often exceed allowable publication budgets or require special permission to be obtained. Many early-career researchers and postgraduate students lack funding for publishing fees altogether. This financial barrier prevents diverse voices from participating in the global research discourse. It disproportionately influences who is cited, reviewed, and noticed, especially by AI systems trained mainly on open-access material. As Cabrerizo[5] notes, even researchers with grants in low-income settings may find APCs prohibitively expensive, emphasising the need for policy reforms to foster more equitable access to publishing opportunities.

Therefore, those able to afford publishing open access enjoy greater visibility (including in AI-curated reviews), while those unable to do so become further marginalised. This economic reality complicates the concept of open source as a universal advantage. APCs have encouraged a pay-to-publish system that upholds existing inequalities under the guise of openness.

Another serious consequence of AI's limited access to scholarly literature is the increased risk of misinformation. With restricted access to the full range of peer-reviewed content, AI tools therefore rely on lower-quality, predatory or non-peer-reviewed sources that are freely available online. An additional limitation of current AI systems is their inability to reliably place research findings within a chronological context (the authors' personal experiences with using ChatGPT4.0). As a result, AI tools may cite studies whose results have been superseded by more recent data or advances in technology, unintentionally promoting outdated or inaccurate information. As Himmelstein and co-authors[3] demonstrate, widespread use of the pirate site Sci-Hub highlights the demand for open access to paywalled literature, emphasising the limitations that AI systems face in accessing such content legally.

If AI-generated content relies on incomplete or poor-quality data sets, it leads to a distorted understanding of complex topics. For example, AI might overlook important nuances in medical, environmental or genetic research because the relevant studies are inaccessible. This becomes especially problematic when AI outputs are accepted uncritically by students, policymakers or other researchers, thus spreading inaccuracies and oversimplifications.

These risks are worsened by the fact that AI systems, unlike trained researchers, often cannot differentiate between historically significant but outdated studies and more recent studies that utilise newer, improved technologies.

Insight: Policy recommendations for South Africa

To minimise inequities in AI-supported scholarship, reforms should concentrate on licensing, affordability, transparency and training:

- National bodies such as SANLiC, NRF and ASSAf should negotiate AI-readable licences as part of subscription and transformative agreements.

This would allow not only institution-bound systems but also public LLMs to access paywalled content under fair-use conditions.

- The NRF and DHET should establish a ring-fenced APC fund for early-career researchers and unfunded scholars.

This would prevent APCs (ZAR40 000–ZAR60 000 per article) from taking a disproportionate share of limited grant funding.

- South African journals, particularly those on the SciELO SA platform or subsidised by ASSAf, should implement disclosure standards for AI use in literature reviews and manuscript preparation.

This would foster trust and encourage critical reflection on how AI impacts research synthesis.

- Universities and professional societies ought to offer training on responsible AI usage in research.

This should include guidance on open access biases in LLMs and strategies for triangulating with subscription databases.

- Developers of academic AI systems should disclose when outputs are based solely on open-access sources.

This transparency would help researchers complement AI outputs with manual searches or institutional access.

Implication: With targeted national action, South Africa can reduce inequities and ensure its researchers stay visible and influential in the AI-driven global scholarly landscape.

Conclusion

Artificial intelligence offers great potential to transform scholarly communication, but its current reliance on open-access content risks worsening existing inequalities in publishing, citation practices and knowledge production. As seen in the South African context, the combination of limited subscription access and high APC costs places local researchers – especially postgraduates and early-career scholars – at a structural disadvantage compared to peers in wealthier regions.

These inequalities are not unavoidable. As outlined in the policy recommendations, South Africa has a clear chance to take proactive actions: negotiating AI-compatible licences through SANLiC, ring-fencing APC support for vulnerable researchers, adopting disclosure norms in national journals, and enhancing training for responsible AI use. By implementing these measures, national bodies (NRF, ASSAf, SANLiC, the Department of Higher Education and Training, and universities) can ensure that South African scholarship remains visible, credible and influential in an era increasingly driven by AI-assisted research.

If left unchecked, the selective visibility of open-access sources may lead to a form of algorithmic canonisation, where the scholarly canon is defined more by machine access than by intellectual merit. Achieving equity for both humans and machines is thus more than just a technical issue – it is a policy choice. Taking action now will promote a more inclusive and representative scholarly record, safeguarding the integrity of knowledge creation in South Africa and across the broader Global South.

Declarations

We have no competing interests to declare. We used ChatGPT 4.0 to correct spelling and grammar and to perform a similarity check. Both authors read and approved the final manuscript.

References

1. Stokel-Walker C, Van Noorden R. What ChatGPT and generative AI mean for science. Nature. 2023;614(7947):214–216. https://doi.org/10.1038/d41586-023-00340-6
2. Van Noorden R. Trouble at the text mine. Nature. 2012;483(7388):134–135. https://doi.org/10.1038/483134a
3. Himmelstein DS, Romero AR, Levernier JG, Munro TA, McLaughlin SR, Greshake Tzovaras B, et al. Sci-Hub provides access to nearly all scholarly literature. eLife. 2018;7, e32822. https://doi.org/10.7554/eLife.32822
4. Klebel T, Ross-Hellauer T. The APC-barrier and its effect on stratification in open access publishing. Quant Sci Stud. 2023;4(1):22–43. https://doi.org/10.1162/qss_a_00245
5. Cabrerizo FM. Open access in low-income countries — open letter on equity. Nature. 2022;605(7911):620. https://doi.org/10.1038/d41586-022-01414-7