Information Research - Vol. 31 No. 2 (2026)

Assessing open access scholarly journals for integration into artificial intelligence research assistants

Sanja Gidakovic, Heather Moulaison-Sandy, and Jenny Bossaller

DOI: https://doi.org/10.47989/ir31263095

Abstract

Introduction. Freely available standalone AI research assistants such as Elicit and Consensus are used by academics to find relevant literature. These systems rely extensively on freely available sources, including open access journal content. No baseline for understanding the level of quality of such journals used in these assistants has been carried out.

Method. A sample of 807 English-language journals from the Directory of Open Access Journals that became open access before 2021 was investigated for quality metrics using SCImago rankings and other defining characteristics and analysed in conjunction with the Directory data.

Analysis. Scimago journal ranking quartile scores were recorded for each of the journals. Descriptive statistics were produced using Excel, and visualizations using Tableau Public.

Results. Of our sample, over half were ranked in Scopus, and many were in quartile 1. Many university or small association journals were unranked.

Conclusions. AI research assistants may miss some high-quality open access content due to reliance on metrics. Commercial enterprises play a large role in sources used to produce content, effectively gatekeeping the process and potentially shaping the creation of new knowledge.

Introduction

Generative AI systems and large language models (LLMs) are widely used by authors of scholarly and scientific publications to help with a variety of tasks, including conceptualization, research, and writing. AI and LLMs are under constant development, with each iteration seeing improvement rapidly. AI has been incorporated into search engines (e.g., Scholar Labs (Google, 2025), released in November 2025) and subscription library databases (e.g., Exlibris Primo (Clarivate, 2025)), enabling researchers to use natural language to search scholarship and generate overviews of the results. Some chat-based GPTs can also suggest better queries, along with other features (So, 2025). All these improvements have the potential to give scholars an edge in writing, in terms of speed, expression, and general productivity (Cruz et al., 2026).

Standalone tools, AI research assistants, incorporate these same functionalities, locating scholarly literature and summarizing and evaluating findings in articles. Faix (2025) defines these systems as tools 'that merge artificial intelligence–driven capabilities with traditional library database functionality, allowing researchers to take advantage of the power of AI to locate articles quickly and efficiently on topics along with providing summaries and evaluations of those articles' (p. 344). Critics of AI use in scholarship often strongly caution users about the "black box" phenomenon since the inner workings of LLMs are not transparent (e.g., the sources are not clear, the accuracy of results cannot be verified, hallucinations appear, etc.) (von Eschenbach, 2021) and LLMs have also been shown to propagate biases (Noble, 2018). One of those problems, citation hallucination, has been largely addressed in AI research assistants (and beyond) with retrieval augmented generation, 'a technique which allows them to retrieve answers from external sources such as web or database searches to improve answer accuracy' (Zhao, 2025, para. 2). This has vastly improved usefulness of AI-powered tools such as AI research assistants for researchers as they gather and summarize relevant literature.

As mentioned, research assistants are integrated in commercial platforms such as Google Scholar and library systems by Clarivate. Standalone AI research assistants such as Elicit, Consensus, and SciSpace, are separate but related tools and are relatively well-known and well-studied (e.g., Cruz et al., 2026). Elicit, Consensus, and SciSpace all search across multiple sources, harvesting from open access databases (Moulaison-Sandy et al., 2025), including Semantic Scholar, OpenAlex, PubMed, and others. These AI research assistants use open access sources, summarizing texts, comparing the findings, etc. Some are beginning to partner with known publishers to enhance their corpora, for instance American Chemical Society Publications, Sage, Wiley, and Taylor & Francis (ACS Publications, 2025; Elechko, 2026). The choice of articles is, in some cases up to the user, with Consensus, Elicit, and SciSpace all allowing users to restrict output to articles that have been published in journals with a SCImago scientific journal rank (SCImago, 2025) (e.g., first. quartile) (see Table 1) to control the relative quality of the items included in their search and resulting summary. Not all open access journals, however, have SCImago rankings, even if scholars in the field deem them to be high quality; this means that not all such journals will be ranked, even if they are good.

AIRA	Where it gets data	How it ranks results
Consensus https://consensus.app/	Searches ~200M peer-reviewed papers from OpenAlex, Semantic Scholar, and Consensu's own web crawl. Source: https://tinyurl.com/5cky9vmm	Consensus retrieves most relevant papers using semantic and keyword matching, and then it re-ranks the top 1,500 papers using recency of publication, citation count, and journal impact / reputation. Consensus provides the SJR quartile ranking filter to limit the search to top quality journals. Source: https://tinyurl.com/bdpj8bpy
Elicit https://elicit.com/	Searches ~ 138 million academic papers from Semantic Scholar, PubMed, and OpenAlex, covering all academic disciplines. Source: https://tinyurl.com/48wbu5hf	The citation data comes from Semantic Scholar. It includes quality signals from citation counts, journal information, methodological details. Elicit currently does not deeply evaluate trustworthiness and acknowledges limitations in judging paper quality. Elicit includes a journal quality filter based on SCImago Journal Rank (SJR) to limit the search to journal quality. Source: https://tinyurl.com/mvvwphw5
SciSpace https://scispace.com/	Searches ~200M papers from OpenAlex, Semantic Scholar, Google Scholar, and other scholarly repositories. Source: https://tinyurl.com/374s2uhw	SciSpace selects top papers by prioritizing relevance and the quality of journal (SJR ranking) or conference they come from using quartile and rating systems rather than citation counts. For conference rating it uses reputation and impact ratings. SciSpace focuses on semantic relevance first, then filters papers from high-quality journals / conferences. Source: https://tinyurl.com/mw6p5n2j

Table 1. Three AI research assistants, their data sources, and their results rankings.

Open access quality control

The Directory of Open Access Journals (DOAJ) offers another quality-control mechanism that is both authoritative (Morrison, 2008; Scholastica, 2018) and inclusive. The Directory was created in 2003 to bolster confidence in the rapidly proliferating open access journal market. To be included in the database, journals must meet defined criteria for openness, publishing, and peer-review standards. While inclusion in the Directory does not guarantee that a journal is high quality, it does show that it has passed the screening process.

AI research assistants seek to connect users with *high-quality* research, but quality indicators for journals have some well-documented flaws, many of which are discussed by Garfield (2006), who first proposed the journal impact factor ranking system in 1955. The ranking system can be gamed through self-citation or 'cartels' seeking to boost their journal's prestige (Kojaku et al., 2021), and outliers (such as a heavily cited article) can skew metrics. Even though it is widely used in measuring scholarly output, Tennant et al. (2019) write that '*Empirical evidence shows that the misuse of the JIF—and journal ranking metrics in general—creates negative consequences for the* *scholarly communication system*' (p. 5). Another criticism is ownership: ranking systems exclusively focus on articles in proprietary databases. Clarivate owns the Web of Science database and the impact fact that is calculated using the platform. In the absence of better, presumably qualitative approaches to evaluating individual articles, metrics for journals are nonetheless frequently used as a proxy for quality despite recent efforts to move away from them toward more qualitative approaches to evaluating (e.g., CNRS, 2025).

Open access in AI research assistants

The nature of the sources returned and summarised by the standalone AI research assistant still depends on several factors, such as whether the user is paying for a premium version, how or whether the user has configured its search capabilities, and what sources are available. Cost continues to be a factor; AI research assistants available on the market do not provide access to full-text articles behind paywalls, though they may locate open access versions if they are available (note that libraries may also incorporate AI research assistants that link to their full-text articles). They also have access to abstracts in Google Scholar. If the user is filtering results by the SCImago ranking, using the quality designation given by the assistant), pre-publications in a repository would presumably be filtered out. We do not know if any AI research assistants mine pirate databases, such as SciHub, though there is evidence that some AI providers are accessing as much content as possible, both legally and otherwise (Ridenour et al., 2025).

The intersection of AI (especially generative AI) and open access has not been well explored, even if the two are mutually reinforcing. LLMs rely on text, and much of the text used in generative pre-trained Transformers (Jones, 2024; Villalobos et al., 2024) as well as AI research assistants (e.g., Moulaison-Sandy et al., 2025) is open access. How well does open access serve AI research assistants? How well do quality indicators for these assistants serve scholars?

Research questions

To probe these questions and to address the gap in the literature, we propose a study to address the following research questions:

> RQ1: Of a sample of open access journals listed in the Directory of Open Access Journals that are likely harvested by standalone AI research assistants, what can be ascertained about their SCImago rankings?
>
> RQ2: Based on DOAJ listings of open access journals and their attributes (e.g., article processing charges required, subject matter, place of publication, publisher), what can be said about the extent of coverage that is potentially available to standalone AI research assistants via such journal articles?

This research fills a gap in the literature on the relationship between gatekeeping in scholarly research (by SCImago quality ranking) and scholars' search behaviour using AI research assistants. AI research assistants (Consensus, Elicit) primarily use SCImago's quartile ranking filter to help users limit their search to top-ranked ("high quality") research. To examine this gap, we compiled journal titles in the Directory that are ranked in SCImago, which also reveals which journals are left out of search results.

Review of the literature

Standalone AI research assistants primarily work through open access sources; thus, the quality of content produced depends on open access journals and other such sources. Because there are so many open access publications, researchers need a way to filter their results by quality. This literature review will describe problems with the way that AI research assistants have chosen to define high-quality research by describing (1) open access and its quality control, (2) impact factor and SCImago rankings, and (3) AI research assistants, especially in relation to access and quality control.

Open access journals in the scholarly publishing ecosystem

Open access journals were created to provide free access to scholarly research publications without payment or subscription barriers for readers, thereby equalising access to scholarship. Publishing and hosting a journal, though, is not free, and the open access model has evolved. *Green* open access allows authors to self-archive a version of a paper in a paywalled journal, while *gold* open access requires authors to pay an article processing charge (Open Access Network, 2026). Journals that do not charge authors to publish do so under a model termed *diamond open access*. The maintenance of diamond journals is usually carried out by non-commercial, non-profit organizations, associations, or academic institutions (Fuchs & Sandoval, 2013).

Some traditional (paywalled) publishers offer authors the option to publish open access for a fee, but there are also very large publishing houses that were established specifically to publish open access journals. Their emergence brought (sometimes deserved) criticism due to unscrupulous business practices (e.g., high processing charges and no peer review), and those became known as "predatory journals," which were tracked by Jeffrey Beall, a librarian who maintained a watchlist between 2008 and 2024, though the Website is still available (Krawczyk & Kulczycki, 2021). The list was criticized sharply for its cultural biases, among other things (Yeates, 2017), but open access still carries a negative connotation among some researchers. Kumari and Subaveerapandiyan (2025) found that there are differences in authors' perceptions of the visibility and equity of open access models. For instance, authors tend to favour gold open access for visibility and potential citation increase, but they expressed concerns about the cost of processing charges. This was especially pronounced among respondents in developing countries. There were also concerns regarding the sustainability of diamond open access, even though it was perceived to be more inclusive because of its no-cost structure.

The Directory of Open Access Journals (DOAJ) was created as a quality-control mechanism for OA journals. DOAJ vets OA journals based on stated principles of transparency and best practice (https://doaj.org/apply/transparency/) such as DOIs for articles, a published editorial board, peer review, and others. Research has found that OA journals are comparable in quality to subscription journals in most fields, especially in the health sciences (Björk & Solomon, 2012; Pastorino et al., 2016). DOAJ only indexes gold and diamond OA journals.

The Directory also improves the discoverability of journals. Discoverability refers to how easily an article can be found through common pathways researchers use to find articles, such as library databases, Google Scholar, PubMed, and others. Even when users rely on general search engines like Google, they can usually find the full text of open access articles. However, the discoverability of articles varies across different models. Dote Pardo (2025) concluded that open access does increase publications' visibility, but 'Factors such as modality (green [i.e., self-archiving] vs. gold), repository infrastructure, journal quality, index coverage, international collaboration, and the availability of multiple copies decisively influence the magnitude of this effect' (p. 12). Thus, while free access to research is always a bonus, other factors, such as indexing and repository infrastructure, impact how these articles are discovered.

Open access publishing is also uneven due to open-access mandates for government-funded research. That led to growth in associated journals (Eve, 2017). There are more humanities open access journals now than in the past, but most are not open access. Note that publishing is very different in the humanities; research rarely requires government funding, and the journals are not exorbitantly priced.

SCImago journal ranking quality metrics

The SCImago journal ranking system is publicly available and provides information on journals indexed in Scopus, a citation and abstract database that covers literature across many disciplines. The system was inspired by Google's PageRank and was developed by a team of researchers at Elsevier and the SCImago Research Group to quantify journal prestige based on the prestige of the citations rather than on simple citation counts (Gonzalez-Pereira et al., 2009; Colledge et al., 2010). Leydesdorff (2009, p. 1328) wrote that the system is the Scopus equivalent of the Web of Science journal impact factor (albeit with a more complicated computation). The system's quartiles are calculated using three years of citations received for articles in each journal, the origin of the citations, and the field in which the journal is situated, among other factors, to arrive at a numerical value representing *prestige* (Gonzalez-Pereira et al., 2009). Journals must be indexed in Scopus to be considered. Scopus tracks citations to individual articles, providing a rich source of data for the journal-level metrics produced by the system.

Journal rankings are used as a quality indicator, but they are not evenly useful. Thelwall et al. (2023) note that impact factors 'have the advantage that in some fields [e.g., medicine] they seem to broadly reflect quality hierarchies' and that they 'are relatively objective and draw on many individual academic decisions' (p. 3916). In the arts and humanities, though, scholarship 'contribute(s) insights rather than building on hierarchical knowledge, so positive correlations between citation counts and quality are not necessarily expected' (p. 3929). Scopus does cover most academic disciplines, but they are unevenly represented. Studies have found that there is greater representation of natural sciences and engineering than social sciences and arts and humanities in Scopus (Mongeon & Paul-Hus, 2016; Pranckutė, 2021). At the same time, Scopus has wider coverage of non-English literature and social sciences than Web of Science (Pranckutė, 2021). Mongeon and Paul-Hus (2016) report that 'Scopus covers less than 25% of journals in [social sciences and arts and humanities], while Web of Science covers less than 15%' (p. 6).

How open access and DOAJ relate to the SCImago journal ranking

Inclusion in Scopus and in DOAJ are both quality controls, but they work very differently. The Directory has adopted the San Francisco Declaration on Research Assessment and thus does 'does not endorse using Impact Factors or other bibliometrics to assess the quality of journals or researchers' (DOAJ, n.d.). DOAJ indexes all open access journals that apply and meet DOAJ standards (DOAJ, 2026). Inclusion in Scopus requires some elements that are essentially identical to the Directory's inclusion criteria (e.g., a rigorous review process, ISSN registration, statements of ethics), but Scopus adds requirements of English-language abstracts and titles, and statements on publication malpractice to guide the journal in the case of ethical concerns with content (Elsevier, 2026).

Scopus currently indexes approximately 8,700 open access journal titles from a wide range of disciplines. The Directory's journals may be ranked or unranked in systems like SCImago's and the journal impact factor; many are both high impact and highly ranked (i.e., in Quartile 1) (e.g., PLOS Medicine, BMC Medicine, *International Journal of Engineering Research & Technology (IJERT)*, and *Journal of Engineering Science and Technology (JESTEC)*). However, many open access journals are not indexed at all because they are regional, too niche (Veretennik & Yudkevich, 2023), or are too new to have established three years of data needed for the ranking. Such journals may have prestige among a small, specialized group.

AI research assistants

AI's strength is synthesis, not reasoning. Humans are still needed to produce ground-breaking work (e.g., Vallor, 2024). Being able to purposefully and intentionally act on goals, or human agency, is key to understanding the promises and limitations of AI (Floridi, 2025). Furthermore, while humans are limited by biology, cultural expectations and beliefs, inefficiencies, and technical capabilities, 'human agency is the only agency that shapes other agencies' (Floridi, 2025, p. 23). Humans are capable of ethical reasoning, and they understand rights and responsibilities. AI exceeds certain human limitations, but it is unable 'to transcend predefined objectives through self-determination' (p. 24). In other words, LLMs are mirrors that reflect what already exists (Vallor, 2024). AI research assistants supporting scholarly inquiry cannot go outside of the bounds of what they are programmed to do and what they have access to. They exist and draw exclusively from a 'box,' (however large or small that box is), and perform within set limitations. However, they are particularly useful in the research process. Schryen et al. (2025) found that AI (including ChatGPT and Google Gemini) can synthesize research, aggregate evidence, criticize prior research, identify research gaps, and even develop a research agenda.

AI research assistants 'function as digital collaborators that can process vast amounts of scientific literature, generate hypotheses, design experiments, analyze data, and even assist in drafting research papers' (So, 2025, p. 1). The Assistant should be able to understand context and personalize information searching and production, serving as a personal research collaborator and thus speeding up the entire research process (So, 2025, p. 4). Van de Schoot et al. (2025) reported that AI-aided search was more thorough than human-alone; their 'machine-aided techniques helped find studies with missing keywords, unusual phrasing, or limited indexing' (p. 1). Similarly, Bernard et al. (2025) emphasise that AI research assistants function best as complementary tools for scoping new topics and identifying papers potentially missed by traditional searching and caution against relying entirely on an AI assistant to find studies. This is often described as a 'human in the loop.' Aside from searching, AI and AI research assistants can improve communication, especially for authors whose native language is not the language of publication.

Standalone AI research assistants like Consensus, Elicit, and SciSpace mine open access databases, including OpenAlex, Semantic Scholar, and Google Scholar, for content, providing users with a vast body of literature. OpenAlex and Semantic Scholar both index over 200 million scholarly documents (Chawla, 2022; Semantic Scholar, 2026). Scopus only recently passed the 100 million item threshold (Feldner, 2025). Maddi et al. (2025) found that OpenAlex indexes more open access journals and is more international than Scopus or Web of Science. Consensus, Elicit, and SciSpace, therefore, have access to more international open access material, but they also give users the option to filter results by the SCImago quartile ranking system (Elechko 2026; *Filter by Journal Quality*, n.d.; *How Top Papers Are Selected in SciSpace*, n.d.). Simrad et al. (2024) analysed open access journals in OpenAlex, Scopus, and the Web of Science by language, field, and country by using the Directory as a reference point. They found that most of the Directory's journals are included in OpenAlex whereas Scopus and Web of Science cover less than half. Further, they found OpenAlex included 'over 12,500 journals indexed, including 60% of all diamond open access journals that are not found in either Web of Science or Scopus' (p. 11). Scopus does not include many prestigious unranked open access articles, or those from newer journals. Thus, by adopting quartile ranking as a measurement of quality, the researcher might miss important publications. See Table 1 for more information about data sources for these three AI research assistants and how results are ranked in each.

There are a few other problems to mention, along with some unknowns. First, the wider representation of the natural sciences in open access publishing means that AI research assistants also favour the sciences. Alison Faix (2025) writes that 'Consensus best covers topics in the sciences and to a lesser extent covers topics in the social sciences. Consensus does not cover the arts or humanities disciplines' (p. 345). Elicit, on the other hand, has a more even coverage of social sciences, but still is not as strong in retrieving humanities research. A smaller or niche AI research assistant could be more useful than a generalist one (Ab, 2024), For instance, a humanities scholar would find an AI assistant trained in humanities literature more useful than a multidisciplinary one that favours the natural sciences by default.

A second problem is that some AI research assistants have problems with incorporating certain file formats. For example, Consensus uses its own Web crawlers to find scholarly content, and these perform better with HTML versions of articles than with PDFs (Montague-Hellen, 2024), even though PDFs are a common format for scholarly publishing. Thus, some of the challenges also stem from how outputs are formatted and not only from open access availability.

Regarding unknowns, it is unclear exactly what AI research assistants can crawl. While we know that they mine open access databases, there are other sources. Some authors deposit preprints of their articles in a repository with a DOI, (thus making the content available to AI), but other publishers and authors restrict AI from accessing and using their content. As such, certain content will not be summarized by the database's AI assistant due to restrictions imposed by the publisher or author. Should that become an issue for AI research assistants that are primarily based on open access content, scholars are already urging for increased publishing in open access and to specifically add licenses to scholarly output that allow for training in AI in order to include high-quality content that will influence how AI models are trained (Montague-Hellen, 2024).

Method

To understand the relationship between open access journals and quality rankings using SCImago's ranking, since such journals are sources used extensively by standalone AI research assistants and the SCImago system is a quality indicator used in the three AI research assistants considered, the DOAJ CSV file of journals was downloaded from the DOAJ site (https://tinyurl.com/5yvtpuxa) on October 23, 2025. The DOAJ file contains information about each of the 22,168 open access journals it indexes, including URLs, processing charges, information about country of publisher, name of publisher, ISSN of journal, language of journal, subjects covered, and much more.

For this project, journals that publish exclusively in English (N=807) were retained in alignment with notions of English being the lingua franca of scholarship (e.g., Pradier et al., 2026). To focus on journals with the required three years of data to qualify them for inclusion in SJR, journals that started to publish all content using an open license in 2021 or after were excluded.

Because manual analysis was needed, a smaller set of 807 articles was established through systematic sampling, with every 10th journal retained for further analysis against SJR-ranked journals using the =MOD(ROW(), 10) function in Excel and retaining journals in row numbers ending in "0." Over a two-week period, members of the research team looked up each journal in SCImago's public interface, recording the ranking quartile in a spreadsheet. Team members verified each other's entries. For journals without quartile information, 0 for "unlisted" was recorded. Analysis was performed in spreadsheets and visualizations (e.g., tables) were created using word processing software and with Tableau Public visualization software (https://public.tableau.com/).

Results

Scopus indicates on its source page that it indexes 8,700 open access journals. The DOAJ indicates it indexes approximately 22,168 journals, a much larger number. The current sample of DOAJ journals includes only English-language journals published prior to 2021 that have had sufficient time to earn a quartile ranking through SCImago's system, given the need for three years of Scopus citation data.

SCImago rank

A majority (n=475, 59%) of the open access journal titles from the sample were ranked in SCImago. The remainder (n=332, 41%) or third quartile. Relatively few were given the lowest rank of Q4 (n=47; 6%). See Table 2.

SJR Rank	Count	Percentage
1	157	19.45%
2	143	17.72%
3	128	15.86%
4	47	5.82%
Unlisted	332	41.14%
Total	807	100%

Table 2. SCImago rank for journals in the sample (N=807).

Initial publication date and SCImago ranking

Longevity of a journal did not improve its ranking. In other words, ranks were not necessarily higher for journals that had been published longer as open access. In fact, none of the journals published before 1990 were Q1. Of the journals first published open access between 2010 and 2020, each year more were unlisted than ranked as Q1. See Figure 1.

When did the journal start to publish all content using an open license?

SJR 0,..	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Q1	5	8	6	4	12	6	11	16	11	7	8	10	11	20
Q2	5	3	6	4	3	8	16	8	12	11	8	8	13	15
Q3	3	8	11	12	7	8	5	8	11	7	10	7	3	4
Q4	2	2	1	4		2	2	4	3	3	4	4	2	4
Unlisted	8	8	9	13	17	23	24	23	24	32	29	22	36	30

Figure 1. Start date when journals were first published open access, by SCImago ranking. See online version in Tableau (https://bit.ly/3OOnWbR) or archived version at the Internet Archive (https://archive.org/details/httpspublic.tableau.comappprofileheather8449vizoainsjropenlicense) (https://bit.ly/47qJHom) for more detail.

Article processing chares and SCImago ranking

About half the DOAJ journals sampled charged for processing, as gold open access. Of the journals that did not charge (i.e., diamond open access; indicated in Figure 1 on the left), over 200 were not included in Scopus or otherwise did not have a quartile assigned. Of the journals that did charge, the majority of ranked journals were Q1 or Q2, suggesting that charges help support "prestige" or perceptions of quality in some way. See Figure 2.

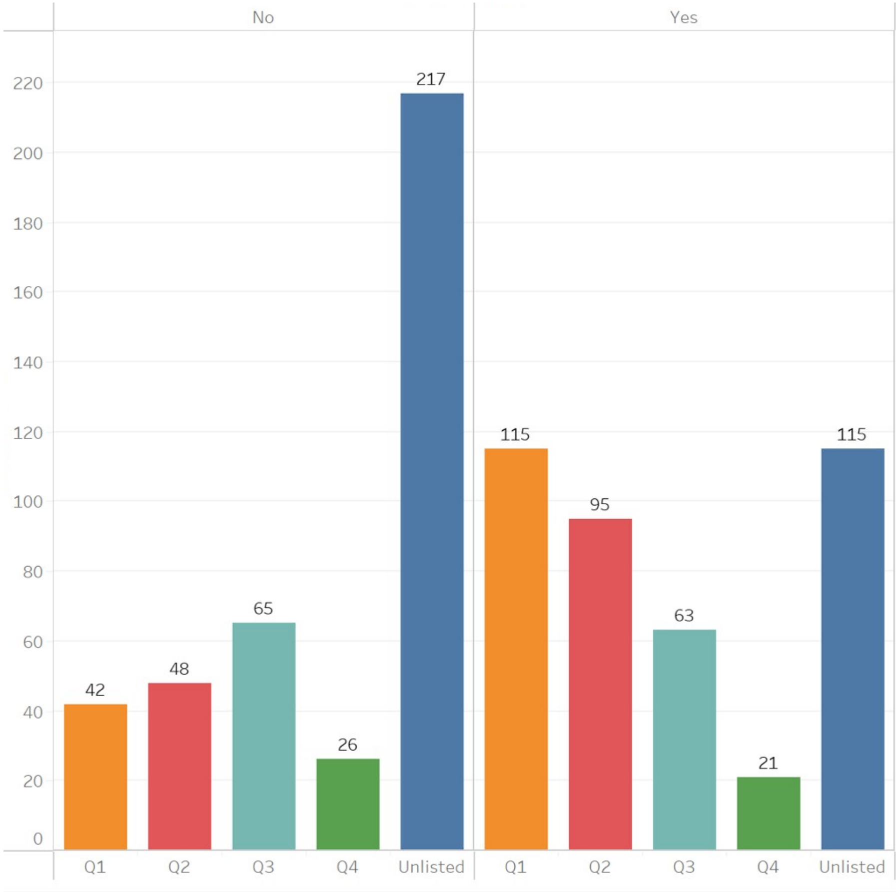

Figure 2. Number of journals in the sample (N=807) that did not ("No"; i.e., diamond) and that did ("Yes"; i.e., gold) charge for processing by SCImago quartiles.

Journal subject and SCImago ranking

The subject of the journals in the sample, as presented in the DOAJ data, ran the gamut. Many focused on medicine, science, and technology. Some journals were generalist or focusing on the social sciences, with a smaller number representing the humanities. See Figure 3.

	SJR 0,1,2,3,4				
Subjects	Q1	Q2	Q3	Q4	Unlisted
Agriculture	3	1			4
Agriculture | Science: Biol..			2		
Agriculture | Science: Bot..					1
Agriculture | Technology					1
Agriculture: Agriculture (..	1	1	2		2
Agriculture: Agriculture (..				1	
Agriculture: Agriculture (..		1			
Agriculture: Animal culture			1		
Agriculture: Animal cultur..					1
Agriculture: Animal cultur..					1
Agriculture: Animal cultur..	1				
Agriculture: Animal cultur..	1		3		2
Agriculture: Aquaculture. ..			1		
Agriculture: Aquaculture. ..					1
Agriculture: Aquaculture. ..	1				
Agriculture: Forestry			1	1	1
Agriculture: Plant culture	1	3	1		
Auxiliary sciences of histo..					3
Bibliography. Library scie..					2
Bibliography. Library scie..					1

Figure 3. DOAJ journal subjects by SCImago ranking. See online version (http://bit.ly/47qGyVA) or archived at the Internet Archive (https://archive.org/details/httpspublic.tableau.comappprofileheather8449vizoainsjrsubjectxapc) (https://tinyurl.com/57uuvdbs).

Publisher information and SCImago ranking

Publishers of English-language open access journals are registered in countries across the world, both in English-speaking countries and beyond. Although this is not indicative of the geographic location of the articles' authors, it does provide insight into the international nature of open access publishing and the dominance of English for scholarly journals. See Figure 4.

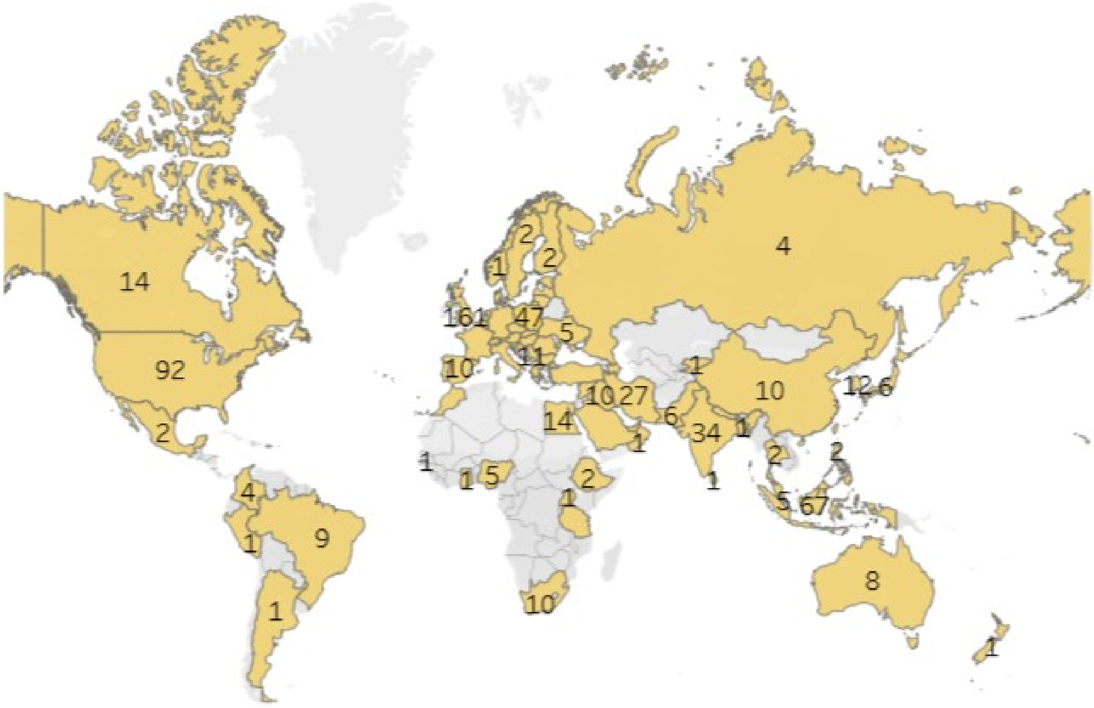

Figure 4. Countries in which sample open access journal publishers are registered.

The sample included for-profit and association journals and those affiliated with a higher education institution, as evidenced by the publisher names. The United Kingdom and United States account for a high number of journals with high SJR rankings. It is worthwhile noting that of the total count of journals for the United Kingdom (n=161) listed in our sample, a large number are of the highest rank (Q1) but many (24%) have no ranking; for the United States, which also has a large number of journals in the sample, more journals are unranked (n=36) than Q1 (n=25). A list of publisher names, by country of publication, showing the SJR ranking of their journals, is available at Figure 5.

Publisher	Country of publisher	Q1	Q2	Q3	Q4	Unlisted
University of Å½ilina	Slovakia			Q3		
University of Al-Qadisiyah	Iraq			Q3		
University of Alberta	Canada					Unlisted
University of Antwerp	Belgium	Q1				
University of Baghdad	Iraq					Unlisted
University of Baghdad, Co..	Iraq		Q2			
University of Belgrade - F..	Serbia		Q2			
University of Belgrade, Te..	Serbia			Q3		
University of Belgrade, U..	Serbia			Q3		
University of Bologna	Italy				Q4	
University of Brawijaya	Indonesia			Q3		
University of Buckingham ..	United Kingdom					Unlisted
University of Canterbury	New Zealand					Unlisted
University of Cordoba	Spain					Unlisted
University of Edinburgh	United Kingdom					Unlisted
University of Garmian	Iraq				Q4	
University of Ghana	Ghana					Unlisted
University of Groningen P..	Netherlands					Unlisted
University of Huddersfiel..	United Kingdom					Unlisted
University of Kufa	Iraq					Unlisted
University of Kuningan an..	Indonesia					Unlisted
University of KwaZulu-nat..	South Africa					Unlisted

Figure 5. Publisher names and countries by SJR ranking of journals. See online version for the full list in Tableau (https://bit.ly/4aSgvsQ) or archived in the Internet Archive (https://archive.org/details/httpspublic.tableau.comappprofileheather8449vizoainsjrsheet10) (https://tinyurl.com/5n96wap3).

Discussion

This study sought to determine what can be ascertained about the SCImago rankings of journals that are included in DOAJ, with SCImago-ranked journals being among the open access sources implied to be high quality. Are these sources representative of the range of perspectives that support the creation of new knowledge? What might be missing from the rankings, and by extension, excluded from consideration by naïve and/or cautious searchers in AI research assistants, given the exclusive nature of the rankings?

Our systematic sample of English-language journals listed in DOAJ, analysed in conjunction with data provided by the DOAJ and through the analysis of the SCImago website found that almost 60% of the sample's titles were listed, and that of these, about 37% were Q1 or Q2 (highly ranked) journals. Those titles would be privileged by AI research assistants when the author asks for highly ranked research to support their own work. This work focuses exclusively on the rankings, rather than attempting to compare them to citation counts (e.g., Walters, 2025), as the ranking already includes elements related to the prestige of citations counts (Gonzalez-Pereira et al., 2009; Colledge et al., 2010).

However, open access journals not highly ranked are worth considering. Over 40% of DOAJ journals sampled were not listed in SCImago's rankings, meaning they had not been selected for inclusion in Scopus. As Van de Schoot et al. (2025) found, artificial intelligence and AI research assistants by extension can expand researchers' vocabulary to uncover new articles and new approaches. However, if the researcher defers judgement to the research assistant to determine quality, and excludes unranked journal articles, they will artificially limit their access to research. Further, Veretennik and Yudkevich (2023) explain that journal-level rankings (impact factor, quartiles, and journal lists) ‘fit the requirements of objectivity and consistency of journal quality assessment’ (p. 3676) but they skew against regional journals in that they ‘reinforce the Matthew effect’ (p. 3676), wherein high IF journals maintain their prestige because of current prestige. Similarly, Thelwall et al. (2023) refer to the ‘positive feedback loop’ of the journal impact factor.

Although longevity of a journal might informally be associated with quality, in this sample, open access journal age was not a positive indicator for SCImago ranking (see Figure 1). Older open access journals were not ranked in the highest quartile. Although perhaps encouraging for younger fields, if older journals are not favoured in SCImago rankings, it could mean that researchers relying on Q1 sources in standalone AI research assistants will miss seminal evidence. If standalone research assistant systems are configured based on current journal quartile, that will also keep highly ranked older work from being considered, depending on how the systems are configured.

Topicality of open access journals in the sample was difficult to quantify given the broad range of headings and subheadings assigned in DOAJ to journals. Further analysis should clarify whether certain subjects are favoured in SCImago. This is because some AI research assistants favour science, technology, engineering, and medicine (STEM). Already, there is more open access in STEM because of funding requirements for publication (Eve, 2017). Fewer humanists may be compelled by funders to publish their work open access. Figure 3 provides insight into the range of subjects and their associated SCImago rank, with evidence that the sciences are better addressed in ranked open access journals than the social sciences or humanities (Faix, 2025), potentially leaving specialists in fields outside of science to have reduced access to content.

The literature confirms that open access journals that do not make a processing charge (i.e., diamond open access) best support publishing opportunities for a broad range of researchers and scholars who do not have grant support for publishing (Kumari and Subaveerapandiyan, 2025). These journals emerge as being highly ranked, which is encouraging. Specifically, 181 diamond open access titles were ranked, with n=90 being highly ranked (i.e., Q1 or Q2). In comparison, gold open access includes 294 journal titles that were ranked, with n=210 as highly ranked (Figure 2). Although gold open access is less fundamentally accessible for authors than diamond open access, gold open access does still produce consistently ranked journals.

The investigation of publisher information and SCImago ranking indicates that the content of the journals may be less a factor than their publisher’s status for inclusion in Scopus. Many publishers that include a society or “University” in their name were lower-ranked or not included in Scopus at all. Although the research questions do not probe the reasons, an examination of the publisher information provided by the DOAJ suggests that these journals may have less funding, more limited infrastructure, or a narrower readership (Veretennik & Yudkevich, 2025) than journals from commercial open access publishers. Journals with limited infrastructure are not necessarily optimised for inclusion in Scopus, potentially lacking the required publication ethics and publication malpractice statements. For this, they will provide a venue for local or member authors’ scholarly output. Despite the overarching similarity of Scopus’s inclusion criteria for journals, Scopus is essentially serving a *de facto* gatekeeping function, deciding which journals will eventually have a quartile ranking and the potential to serve as “high quality” fodder for AI research assistants.

Commercial interests and open access

Although the SCImago quartile filter may be useful to researchers for identifying journals that have been cited in prestigious publications (Gonzalez-Pereira et al., 2009), it directly impacts access to the world's open access knowledge. Of the 807 DOAJ journals sampled, 41% (n=332) (Figure 1) were not indexed in Scopus even though they met the criteria for inclusion in DOAJ. Unranked journals emerge as a kind of wild card, as their potential for "quality" is wholly untested other than confirmation that the journal's basic infrastructure is intact, supporting inclusion in DOAJ. Unranked articles may be overlooked simply because of their exclusion from the Scopus ranking system. Making decisions about the potential of the world's scholarship is an awesome responsibility for a single for-profit company, RELX the parent company of Scopus, and a disproportionate one as it relates to discovery and use of scholarly information.

Further, standalone AI research assistants seem to gatekeep as well through their functionalities, limiting to SCImago-ranked journals or open access articles from highly ranked sources. Findings from this analysis indicate that the journals ranked in quartile 1 are typically published by the dominant commercial publishers (e.g., Elsevier, Wiley, Springer, BMC), through open access mega-journals like SpringerOpen, and by universities or societies. Journals not ranked in SCImago were typically published by academic institutions (see Figure 5 and accompanying online table). The country of publication differed between journal with SJR ranks and those without, and many unranked journals were from certain countries, such as Indonesia, India, etc. Small and independent publishers like universities and scholarly societies seem to be at a disadvantage, victims perhaps of the Matthew effect (Veretennik and Yudkevich, 2023) since they do not benefit from the required extant infrastructure.

Limitations

Due to the need for manual analysis as part of the methodology, this research only considered a sample of English-language open access journals indexed in the DOAJ. Automated means of analysing the full set of data could yield different result. Further, this project investigated a sample of open access journals published in English and entirely excluded from consideration multilingual open access journals that might have met Scopus's inclusion criteria. Further, although we did not try to quantify or evaluate 'quality,' notions of quality are inherent in design decisions observed in the AI research assistants discussed, in which rank serves as a proxy for quality.

Implications and future research

Universities, libraries, and individual scholars are investigating and investing in AI agents. Furthermore, AI is being embedded in many products that libraries subscribe to, making this a highly volatile and competitive environment. Both scholars and librarians need to understand the purposes and limitations, strengths and weaknesses, of various AI agents so that they know which ones will best meet their needs, which ones are duplicative (drawing on the same data) as well as the ones to avoid. Based on the quality of the sources, our results affirm that AI research assistants are useful tools at the beginning of the process of finding scholarly literature (Faix, 2025), since they use scholarly data from various open access scholarly databases (Semantic Scholar, OpenAlex) which have been proven to be more representative of scholarly content compared to other databases (Maddi et al., 2025). They are also an option when access to paywalled databases is not available due to their free version user access option.

Complexity abounds. On one hand, open access journals that wish to maintain their ranking might have more guardrails against 'bad' scholarship (e.g., plagiarized, AI-generated, or sloppy papers). On the other hand, standalone AI research assistants feature the ability to limit by SCImago rank, a functionality that gives preference to articles that are indexed in a particular ranking system so that these for-profit companies will do well. If authors use AI research assistants as research partners, they must know the pool of articles they are pulling from, be able to control the quality indicators, and cast a wide net. Increased transparency inspires trust in the systems, and educating users is key.

Future possible avenues for research include working with a larger sample or adopting computerized means of analysis of the full set of DOAJ journals. Additionally, sampling other languages in a follow-up study as a point of comparison and as a way of more robustly understanding not only the role of other languages in standalone AI research assistants, but also the place of English in open access.

Ultimately, at present Scopus and its parent company RELX are serving a gatekeeping capacity to the world's open access literature as it is made available through standalone AI research assistants through its inclusion and exclusion of journals. To adequately accommodate open access sources, Scopus should commit to indexing a fuller range of open access journals. Additionally, AI research assistants have a responsibility. Although they are for-profit companies wishing to legitimize their content maximally, by pretending that rankings are surrogates for quality, they are potentially leaving out high quality sources given the current indexing practices of Scopus. SciSpace admirably promotes understanding 'quality' rather than using the shorthand of quartiles but does not provide a definitive way to implement this notion. Additionally, as AI research assistants become increasingly integrated into academic library databases and into Google Scholar via Scholar Lab, we encourage all of these systems to be ethical stewards the world's knowledge, presenting content fairly and without bias.

The technologies, legalities, and scholarly publishing landscape is in a state of constant flux, so new issues will undoubtedly arise. However, there are implications for some principles in scholarly communication (e.g., indexing, measures of quality, open access, and trustworthiness) that are ingrained in both technologies and practice and deserve attention. The importance of open access literature to the present and future scholarly communication landscape due to its extensive use in AI research assistants should not be underestimated, as the information science community seeks to address ways to represent and include the broadest possible range of scholars and their work.

About the authors

Sanja Gidakovic, Universiy of Missouri: Columbia. sg6kn@umsystem.edu

Heather Moulaison-Sandy, University of Missouri: Columbia. moulaisonhe@missouri.edu

Jenny Bossaller, University of Missouri: Columbia. bossallerj@missouri.edu

References

ACS Publications. (2025, November 18). ACS *Publications partners with Consensus to make scientific literature more accessible.* ACS Publications Chemistry Blog. bit.ly/4b3kjWG

Bernard, N., Sagawa Jr, Y., Bier, N., Lihoreau, T., Pazart, L., & Tannou, T. (2025). Using artificial intelligence for systematic review: the example of elicit. BMC medical research methodology, 25(1), 75. https://doi.org/10.1186/s12874-025-02528-y

Björk, B. C., & Solomon, D. (2012). Open access versus subscription journals: a comparison of scientific impact. *BMC Medicine*, 10(1), 73. https://doi.org/10.1186/1741-7015-10-7

Chawla, D. S. (2022). Massive open index of scholarly papers launches. *Nature.* https://doi.org/10.1038/d41586-022-00138-y

Clarivate. (2025). *Generative AI and the future of library services: opportunities and challenges in a changing technology environment.* https://tinyurl.com/5n78f86n

Colledge, L., de Moya-Anegón, F., Guerrero-Bote, V., López-Illescas, C., & Moed, H. (2010). SJR and SNIP: two new journal metrics in Elsevier's Scopus. *Serials*, 23(3), 215-221. https://doi.org/10.1629/23215

CNRS. (2025, December 1). *CNRS is breaking free from the Web of Science*. CNRS. https://tinyurl.com/bduy7yym

DOAJ (2025, Jan 16). *Inside the Quality Team – Part 1: how we preserve quality at DOAJ. DOAJ Blog*. https://bit.ly/4cw65jJ

DOAJ. (2026). *Guide to applying*. https://bit.ly/3NdVVK9

DOAJ. (n.d.). *Quality control process*. https://bit.ly/4lkY3MX

Dote Pardo, J. S. (2025). Impact of open access on academic visibility: a systematic review of the literature. *Journal of Documentation*, 1-27. https://doi.org/10.1108/JD-08-2025-0216

Elechko, A. (2026). *Consensus research database*. Consensus. https://tinyurl.com/bdpj8bpy

Elsevier. (2026). *Scopus: Content policy and selection*. https://tinyurl.com/yc3w37z2

Eve, M. P. (2017). Open access publishing models and how open access can work in the humanities. *Bulletin of the Association for Information Science and Technology*, 43(5), 16-20. https://doi.org/10.1002/bul2.2017.1720430505

Faix, A. (2025). Consensus: Using AI to analyze scientific literature. *Library Trends*, 73(3), 344-354. https://doi.org/10.1353/lib.2025.a961198

Feldner, D. (2025, March 3). *Scopus data crosses the 100 million item threshold!* Scopus Blog. https://tinyurl.com/bdcjfnwy

Elicit (n.d.). *Filter by Journal Quality*. https://tinyurl.com/mvvwphw5

Floridi, L. (2024). AI as agency without intelligence: on artificial intelligence as a new form of artificial agency and the multiple realisability of agency thesis. *Available at SSRN*. https://doi.org/10.1007/s13347-025-00858-9

Frontiers Media SA. (2020, July 1). Artificial intelligence to help meet global demand for high-quality, objective peer-review in publishing. Frontiers Science News. https://www.frontiersin.org/news/2020/07/01/artificial-intelligence-peer-review-assistant-aira

Fuchs, C., & Sandoval, M. (2013). The diamond model of open access publishing. Why policy makers, scholars, universities, libraries, labour unions and the publishing world need to take non-commercial, non-profit open access serious. *TripleC: Communication, capitalism & critique*, 11(2), 428-443. https://doi.org/10.31269/vol11iss2pp428-443

Garfield, E. (2006). The history and meaning of the journal impact factor. JAMA 295(1), 90-93. https://doi.org/10.1001/jama.295.1.90

Gonzalez-Pereira, B., Guerrero-Bote, V., & Moya-Anegon, F. (2009). *The SJR indicator: a new indicator of journals' scientific prestige*. arXiv. (No. arXiv:0912.4141) https://doi.org/10.48550/arXiv.0912.4141

Google. (2025, November 18). Scholar Labs: an AI powered scholar search. *Google Scholar Blog*. https://bit.ly/4aYtmIE

Gouzi, F., & Tasovac, T. (2025). Gold, green, diamond: what you should know about Open Access publishing models. *DARIAH-Campus*. https://tinyurl.com/4auemnsd

Jones, N. (2024). The AI revolution is running out of data. What can researchers do? *Nature*, 636(8042), 290–292. https://doi.org/10.1038/d41586-024-03990-2

Kojaku, S., Livan, G., & Masuda, N. (2021). Detecting anomalous citation groups in journal networks. *Scientific Reports*, 11(1), 14524. https://doi.org/10.1038/s41598-021-93572-3

Krawczyk, F., & Kulczycki, E. (2021). How is open access accused of being predatory? The impact of Beall's lists of predatory journals on academic publishing. *The Journal of Academic Librarianship*, 47(2), 102271. https://doi.org/10.1016/j.acalib.2020.102271

Kumari, M., & Subaveerapandiyan, A. (2025). Perceptions of open access publishing: a comparative study of gold and diamond models among global researchers. *Alexandria*, 35(1-2), 55-73. https://doi.org/https://doi.org/10.1177/09557490251335952

Leydesdorff, L. (2009). How are new citation-based journal indicators adding to the bibliometric toolbox? *Journal of the American Society for Information Science and Technology*, 60(7), 1327-1336. https://doi.org/10.1002/asi.21024

Maddi, A., Maisonobe, M., & Boukacem-Zeghmouri, C. (2025). Geographical and disciplinary coverage of open access journals: OpenAlex, Scopus, and WoS. *PLoS One*, 20(4), e0320347. https://doi.org/10.1371/journal.pone.0320347

Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics*, 106(1), 213-228. https://doi.org/10.1007/s11192-015-1765-5

Montague-Hellen, B. (2024). Empowering knowledge through AI: open scholarship proactively supporting well trained generative AI. *Insights*, 37(10), 1-9. https://doi.org/10.1629/uksg.649

Morrison, H. (2008). Directory of open access journals (DOAJ). *The Charleston Advisor*, 18(3), 25-28. https://doi.org/10.14288/1.0107434

Moulaison-Sandy, H., Castaño-Muñoz, W., Ridenour, L., & Adkins, D. (2025). AI literature review systems: an analysis of performance, affordances, and outputs for a complex topic in the social sciences. *Information Research an International Electronic Journal*, 30(iConf), 1244–1252. https://doi.org/10.47989/ir30iConf46906

Noble, S. (2018). *Aalgorithms of oppression: how search engines reinforce racism.* NYU Press. https://doi.org/10.2307/j.ctt1pwt9w5

Open Access Network (2025). *Open Access Green and Gold.* https://tinyurl.com/4uxuuzxx

Pradier, C., Céspedes, L., & Larivière, V. (2026). How multilingual is scholarly communication? Mapping the global distribution of languages in publications and citations. *Journal of the Association for Information Science and Technology*. https://doi.org/10.1002/asi.70055

Pranckutė, R. (2021). Web of Science (WOS) and Scopus: the titans of bibliographic information in today's academic world. *Publications*, 9(1), 12. https://doi.org/10.3390/publications9010012

Pastorino, R., Milovanovic, S., Stojanovic, J., Efremov, L., Amore, R., & Boccia, S. (2016). Quality assessment of studies published in open access and subscription journals: results of a systematic evaluation. *PLoS One*, 11(5), e0154217. https://doi.org/10.1371/journal.pone.0154217

Ridenour, L., Thach, H., & Knudsen, S. E. (2025). Library Genesis to Llama 3: navigating the waters of scientific integrity, ethics, and the scholarly record. *Proceedings of the Association for Information Science and Technology*, 62(1), 1063–1069. https://doi.org/10.1002/pra2.1340

Scholastica. (n.d.) A *platform for* OA *journal discovery: interview with* DOAJ. https://tinyurl.com/5dzmtws5

Schryen, G., Marrone, M., & Yang, J. (2025). Exploring the scope of generative AI in literature review development. *Electronic Markets*, 35(1), 13. https://doi.org/10.1007/s12525-025-00754-2

SCImago. (2025). *SCImago journal and country rank: About us*. SCImago Research Group. https://tinyurl.com/ypvjdwac

SciSpace (2025). *How Top Papers Are Selected in SciSpace*. https://tinyurl.com/mw6p5n2j

Semantic Scholar. (2026). *About Semantic Scholar*. https://www.semanticscholar.org/about

So, R. (2025). AI-based scientific research assistants. Project-rachel.4open.science. https://tinyurl.com/bdz63u9j

Tennant, J. P., Crane, H., Crick, T., Davila, J., Enkhbayar, A., Havemann, J., Kramer, B., Martin, R., Masuzzo, P., Nobes, A., Rice, C., Rivera-López, B., Ross-Hellauer, T., Sattler, S., Thacker, P. D., & Vanholsbeeck, M. (2019). Ten hot topics around scholarly publishing. *Publications*, 7(2), 34. https://doi.org/10.3390/publications7020034

Thelwall, M., Kousha, K., Makita, M., Abdoli, M., Stuart, E., Wilson, P., & Levitt, J. (2023). In which fields do higher impact journals publish higher quality articles? *Scientometrics*, 128(7), 3915-3933. https://doi.org/10.1007/s11192-023-04735-0

Vallor, S. (2024). *The AI mirror: How to reclaim our humanity in the age of machine thinking*. Oxford University Press. https://doi.org/10.1093/oso/9780197759066.001.0001

van de Schoot, R., Messina Coimbra, B., Evenhuis, T., Lombaers, P., Weijdema, F., de Bruin, L., ... van Zuiden, M. (2025). The hunt for the last relevant paper: blending the best of humans and AI. *European Journal of Psychotraumatology*, 16(1). https://doi.org/10.1080/20008066.2025.2546214

Veretennik, E., & Yudkevich, M. (2023). Inconsistent quality signals: Evidence from the regional journals. Scientometrics, 128(6), 3675-3701. https://doi.org/10.1007/s11192-023-04723-4

Villalobos, P., Ho, A., Sevilla, J., Besiroglu, T., Heim, L., & Hobbhahn, M. (2024). *Will we run out of data? Limits of LLM scaling based on human-generated data* arXiv. (No. arXiv:2211.04325) .https://doi.org/10.48550/arXiv.2211.04325

Von Eschenbach, W. J. (2021). Transparency and the black box problem: why we do not trust AI. *Philosophy & Technology*, 34(4), 1607–1622. https://doi.org/10.1007/s13347-021-00477-0

Walters, W. H. (2025). Gauging scholars' acceptance of open access journals by examining the relationship between perceived quality and citation impact. *Journal of Data & Information Science*, 10(1), 151–166. https://doi.org/10.2478/jdis-2025-0002

Yeates, S. (2017). *After Beall's 'List of predatory publishers': problems with the list and paths forward*. https://informationr.net/ir/22-4/rails/rails1611.html

Zhao, A. (2025). Trust in AI: Evaluating Scite, Elicit, Consensus, and Scopus AI for generating literature reviews. *Research Bridge*. bit.ly/4sqfY7b

Zoccali, C., & Mallamaci, F. (2023). The changing landscape of scientific communication: Open access, predatory journals and the near future. *Journal of Nephrology*, 36(8), 2209-2212. https://doi.org/10.1007/s40620-023-01702-z

Copyright