Information Research Communications, 2027; 4(1):69-76.
https://inforescom.org

Research Article

Responsible Research Assessment in the Age of Artificial Intelligence: Challenges and Future Directions

Abhay Maurya

Department of Library and Information Science, Apex University, Jaipur, Rajasthan, INDIA.

ABSTRACT

Background: The increasing use of bibliometric indicators, including Journal Impact Factor, h-index, CiteScore, and Field-Weighted Citation Impact, has transformed research assessment by emphasizing measurable research impact. However, their inappropriate or excessive use may contribute to disciplinary bias, research misconduct, and undesirable changes in scholarly behavior. The emergence of Artificial Intelligence (AI) further introduces both opportunities and challenges for responsible research assessment and academic publishing. **Objectives:** This paper aims to critically examine the evolving landscape of research assessment in the context of AI, identify limitations associated with conventional bibliometric indicators, and explore approaches for developing transparent, multidimensional, and responsible assessment systems. **Methodology:** A critical narrative analysis was undertaken of contemporary approaches to research assessment and the emerging role of AI in scholarly communication. Major responsible research assessment initiatives, including the Leiden Manifesto, San Francisco Declaration on Research Assessment (DORA), Coalition for Advancing Research Assessment (CoARA), and Metric Tide, were examined to identify principles relevant to responsible use of quantitative and qualitative indicators. **Results:** The analysis highlights that bibliometric indicators can provide valuable evidence of research influence but are insufficient as standalone measures of research quality. Their misuse may reinforce disciplinary disparities and incentivize metric-driven scholarly behavior. AI offers opportunities for efficient analysis, contextual interpretation, and research evaluation, but also raises concerns regarding AI-generated manuscripts, fabricated references, hallucinated citations, transparency, and research integrity. Integrating quantitative indicators with expert peer review, qualitative assessment, and contextual interpretation provides a more balanced approach to evaluating research contributions. **Conclusion:** Responsible research assessment in the AI era requires a shift from metric-centered evaluation toward transparent, multidimensional, and context-sensitive approaches. Combining appropriate bibliometric indicators with qualitative expert judgment and responsible AI practices can promote research integrity, fairness, and excellence while reducing the risks associated with inappropriate metric use and AI-enabled scholarly misconduct.

Keywords: Artificial intelligence, Bibliometric indicators, CoARA, DORA, Leiden Manifesto, Open Science, Research assessment, Research evaluation, Research integrity, Responsible bibliometrics, Responsible metrics.

Correspondence:
Dr. Abhay Maurya
Department of Library and Information Science, Apex University, Jaipur, Rajasthan, INDIA.
Email: abhaymaurya17@hotmail.com
ORCID: 0000-0002-6097-3619

Received: 21-08-2026;
Revised: 02-09-2026;
Accepted: 19-10-2026.

INTRODUCTION

The assessment of scientific performance has become important in contemporary research policy, governance of higher education, and allocating fund. The expansions of research systems have increased the need of objective and scalable methods to assess impact, influence and quality of scientific studies. The originality and methodological rigor of peer review has made it a cornerstone of research assessment. However, factors like exponential growth in scientific publications and increasing complexity of research ecosystems have rendered exclusive dependence on qualitative assessment difficult. Hence, quantitative indicators derived from bibliometric analysis have emerged as important tools for assessing productivity and impact of research (Aksnes *et al*., 2019; Bhattacharya, 2026).

Bibliometrics is a way of quantitatively assessing publications and the citations to analyze development and trends in scientific studies. Post development of citation indexing by Eugene Garfield, citation analysis has been a central part of assessment in modern bibliometrics (Garfield, 1972). The widespread acceptance of bibliometrics has enhanced the importance of research assessment. Research assessment depends on several key bibliometric indicators. These include JIF, h-index, CiteScore, FWCI, SNIP, and SJR. These indicators are increasingly used in

DOI: 10.5530/irc.4.1.6

Copyright Information :
Copyright Author (s) 2027 Distributed under Creative Commons CC-BY 4.0

Publishing Partner : Manuscript Technomedia. [www.mstechnomedia.com]

academic promotions, allocating funds, institutional rankings, and making science policies. Indicators of bibliometrics have gained widespread acceptance in practices of assessment due to the systematic, explicit, and comparatively objective assessment of research (Hicks *et al*., 2015).

While bibliometric indicators have been widely implemented, their misuse and over scrutiny have prompted criticism. Citation-based indicators were intended to complement expert assessment, not replace it. However, institutions have replaced robust assessments of research quality with numerical indicators, irrespective of the research discipline, the impact of research, the methods employed, and other significant factors. This manifests Goodhart's Law. As bibliometric indicators are used for recruiting, promoting, tenuring staff, and allocating research grants, it is expected that researchers will focus on maximizing indicators instead of their actual contribution to science. The effects of this focus on bibliometric indicators have led to self-citation, citation cartels, salami slicing, pay to publish type predatory publishing, and other practices which are detrimental to research.

AI systems like AI and LLMs have added significant challenges to assessment of research. Rapid development of AI has enabled researchers in literature reviews and writing and editing. However, AI generated drafts, filled links, hallucinated citations, paper mills, and automatic content generation is undermining bibliometric databases and the integrity of scientific records (Walters and Wilder, 2023). The growing popularity of AI in scholarly publishing requires mechanisms that support openness, accountability, and responsible assessment. On the contrary, AI helps detect citation manipulation, publication fraud, and paper mills, besides improving processes of quality assurance of scholarly publications (Bhattacharya, 2026; Notman, 2025).

These changes have led to the emergence of a global movement for responsible research assessment. The international manifestos like DORA (DORA, 2013), CoARA (Hicks *et al*., 2015) and the Metric Tide report (Wilsdon *et al*., 2015) claim that quantitative indicators should be used to complement qualitative expert judgment and not replace it. The latter documents emphasize the strengths of transparency, diversity, contextualization, and responsible use of bibliometric indicators and suggest that an evaluation of individual researchers shouldn't depend on single measures. Recent papers expand these proposals, calling for narrative bibliometrics, multi-dimensional assessment systems, open science practices, assessing societal impact, or providing context-based research evaluation fitting the demands of increasingly interdisciplinary and cooperative science environments (Daraio *et al*., 2025; Kar, 2026; Torres-Salinas *et al*., 2024).

According to recent literature, the application of bibliometrics must be conceived as an ecosystem for enabling transparent, unbiased, and fair evaluation systems. Institutions are invited to consider the bibliometric evidence in conjunction with other sources such as peer review, descriptive accounts, societal impacts, open science initiatives, indicators on research integrity, and subject-specific influences in order to create integrated visions for research assessment excellence (Bhattacharya, 2026; Rushforth and Hammarfelt, 2023). Such multiple dimensional assessment frameworks take into account the insufficient nature of numerical value to capture the excellence of scientific work, and recognize that the metric evidence must be complemented with qualified expert judgment.

This article will review current developments in the field of responsible research assessment against the background of the rise of Artificial Intelligence (AI). It will critically analyze the history and state of art of bibliometric measures, problems with metric based research assessment, chances and threats of AI on scientific communication, and current attempts to introduce globally established systems of responsible bibliometrics. Finally, it will sketch some perspectives to establish transparent, ethic based, context sensitive evaluation system to strengthen research quality and integrity in an AI assisted science.

EVOLUTION OF RESEARCH ASSESSMENT

The last century has witnessed the evolution of research assessment from qualitative expert assessment to complicated, quantitative and multi-dimensional assessment. This reflects the growing volume of scientific output, globalization of research, and increasing demand for transparent and evidence-based assessment systems (Aksnes *et al*., 2019). Opine that while peer review remains the foundation of scholarly assessment, different bibliometric indicators have gained important in measuring performance of research.

Peer review, which depends on different experts who assess the novelty, methodological rigor, implication, and validity of scholarly work, is regarded as the benchmark for assessing quality of research. It provides judgments on theoretical novelty, appropriateness of methodology, and shared importance that are beyond the purview of quantitative metrics. However, the process is criticized for being lengthy, costly, vulnerable to bias, contradiction, conflicts of interest, and disciplinary subjectivity. These limitations have encouraged search for assessment methods capable of supporting large-scale research assessment (Belter, 2015; Wilsdon *et al*., 2015).

The introduction of citation-based assessment marked a paradigm shift in research assessment. Citation indexing by Garfield enabled the systematic measurement of scholarly influence through citation analysis. This laid the foundation for modern bibliometrics (Garfield, 1972). Citation counts provide an objective and scalable means of identifying influential publications, researchers, and institutions. Over time, citation analysis has expanded beyond simple citation counts to include a wide range of normalized indicators designed to facilitate

cross-disciplinary comparisons and improve the accuracy of research evaluation (Aksnes *et al*., 2019).

The h-index, proposed by Hirsch is among the most widely recognized indicators. This index allows for measuring research productivity and citation impact concurrently. Despite its simplicity and wide acceptance, this index has several disadvantages. It fails to cater to new researchers, has inter-disciplinary variations, and fails to account for highly cited publications (Hirsch, 2005).

Another influential metric is the JIF, which assesses journals. Its widespread misuse in assessing individual researchers has received strong criticism due to the skewed nature of citation distributions are considerable variance across disciplines. Global initiatives like the Leiden Manifesto and DORA discourage the use of JIF as a proxy for research quality (Hicks *et al*., 2015; DORA, 2013).

The limitations of JIF have been addressed by CiteScore, developed by Elsevier. Using the Scopus database, CiteScore calculates average citations over a 4-year publication window. It provides broader journal coverage and greater openness in including documents, making it a useful alternative to assess journals. Also, the FWCI normalizes citation performance by accounting for differences in discipline, publication type, and publication year, thereby enabling fairer cross-disciplinary comparisons. An FWCI value more than one indicates that a publication has been cited more frequently than the global average for comparable publications (Elsevier, 2026).

The extension of scholarly influence beyond academic citations has led to altmetrics gaining importance as a complementary indicator. Using social media posts, policy documents, and other reference management platforms, altmetrics measures online attention. The vulnerability of these indicators to media popularity, platform algorithms, and disciplinary differences make these complement traditional bibliometric measures (Priem *et al*., 2011; Torres-Salinas *et al*., 2024).

The progression of research assessment shows a gradual shift from relying on single assessment method toward integrated frameworks that amalgamate peer review with several quantitative and qualitative indicators. The inability of single metric to effectively capture the different dimensions and impact of research quality has led to development of methods that advocate responsible, transparent, and context-sensitive assessment (Bhattacharya, 2026; Rushforth and Hammarfelt, 2023).

PROBLEMS WITH METRIC-BASED EVALUATION

Bibliometric indicators provide consistent and scalable measures of scholarly performance. Decisions regarding recruitment, promotion, funding, and rankings of institutions depend on citation counts, h-index, and JIF. While these indicators improve openness and comparability, over reliance leads to concerns regarding fairness, integrity, and the quality of scientific output. The inability of single quantitative indicator to adequately capture the multidimensional nature of scientific excellence, has led to scholars advocating that metrics should inform and not determine research assessment (Aksnes *et al*., 2019; Bhattacharya, 2026).

Goodhart's Law

Goodhart's Law criticizes metric-based evaluation. This law states that "when a measure becomes a target, it ceases to be a good measure." Originally intended as descriptive indicators of scholarly influence, bibliometric measures have become performance targets influencing academic careers and institutional success. The dependence of hiring, promotion, and opportunities of funding on citations, publication counts, and journal metrics, makes researchers shift their publication behavior from maximizing scientific contribution to maximizing measurable outputs (Bhattacharya, 2026).

The consequences extend beyond individual researchers to engulf entire research system. Universities seeking higher global rankings encourage publications in high-impact journals and disregard local or community challenges. Funding agencies, which rely on numerical indicators, reward productivity over originality. These agencies discourage interdisciplinary or high-risk research. Consequently, metric driven assessment encourages safe research topics that can generate more citations. Such bias distorts research agendas, narrows down the range of ideas and hinders innovative output (Rushforth and Hammarfelt, 2023). The Leiden Manifesto, along with the DORA proposes an interpretation of bibliometric metrics in a certain context, which, inter alia highlights that quantifiable measures should be informative to rather than substitutive of expert opinion (Hicks *et al*., 2015; DORA, 2013).

Gaming Metrics

The rise in popularity of bibliometric metrics has led to gaming the metrics. Gaming is a practice to artificially inflate bibliometric scores. One common way to do this is self-citation, where individuals frequently cite their previous publications to increase their publication counts. This helps increase their h-index and other bibliometric metrics. While self-citations are necessary to establish continuity in research area, excessive use of self-citations gives a misleading representation of scholarly output (Biagioli and Lippman, 2020).

Another unethical practice involves citation cartels. These are groups of authors, journals, or editors who agree to cite one-another to inflate bibliometric indicators. This citation behavior decreases the credibility of citation-based assessment and creates confusing impressions of scholarly influence. Likewise, some peer reviewers request needless citations to their own work during the process of review, to artificially increase

their citation counts and undermine the integrity of peer review (Bhattacharya, 2026).

The recent and rapid increase in predatory journals, which focus on collecting article-processing charges than on the quality of published scholarship and offer rapid publication rates without effective peer review standards or quality control procedures, exacerbate these issues as many articles published in predatory venues manipulate citation networks to boost citation rate (Walters and Wilder, 2023). The proliferation of AI in academic publishing has now enables creation of AI-generated articles as well as fabricated references. This increases the risk of contamination of bibliometric databases (Walters and Wilder, 2023).

A prevalent form of article gaming involves salami slicing or dividing one large research project into several smaller pieces of work for multiple publications. Although this may meet basic requirements for publication, it results in the fragmentation of scientific evidence, duplicates effort on a global level and increases the load on peer reviewers while providing an artificial picture of one's productivity (Biagioli and Lippman, 2020).

Field Differences

One of the major shortcomings of citation-based evaluations is the high degree of variability in citing behavior across fields. Differences between disciplines and academic areas regarding rates of production, collaborations, document forms and citation life cycles have long been recognized. The medical field for instance, typically has high volumes of papers that attract many citations in relatively short period of time compared to Mathematics, Engineering, Human and Social Sciences where there are fewer publications that accumulate slower citation, low total citation rates and lower total citation rates (Aksnes *et al*., 2019).

A direct comparison of raw counts in different fields might lead to unwarranted conclusions about research quality. Scholarly outputs in several fields and in regional languages are underrepresented in major bibliometric databases like WoS and Scopus. These differences have helped develop normalized indicators like the FWCI, which adjusts citation performance relative to field, publication type, and publication year. Nevertheless, normalization cannot entirely eliminate contextual differences, reinforcing the need for discipline-sensitive and qualitative evaluation approaches (Hicks *et al*., 2015).

Overdependence on Journal Impact Factor

The JIF remains among the most significant and controversial metrics of all bibliometric indicators. Although created with the purpose of being applied to journals, the JIF is now commonly employed to judge research work on an individual researcher and their contributions. The inability of quality of journals to represent scientific quality for individual papers has led to criticism of this method (Garfield, 1972; DORA, 2013).

JIF impacts several decisions

Researchers are assessed depending on the impact factor of journals where they publish. This undermines the originality or social significance of their contributions. While promotion committees equate publication venue with research quality, funding agencies use JIF to screen grant applications. These disadvantage young scholars, scholars that focus on emerging and trans-disciplinary fields, publish on specialized/local journals, aim for a prestige-based promotion instead of focusing on scientific impact and contribution to society (Bhattacharya, 2026; Hicks *et al*., 2015).

Overall, based on the limitations of metric based assessment, bibliometric indicators should be seen under the lens of their social, contextual and disciplinary framework and therefore, quantitative methods and tools need to be combined with qualitative peer review and ethical frameworks, defined as responsible research evaluation.

ARTIFICIAL INTELLIGENCE AND RESEARCH ASSESSMENT

AI has emerged as the most transformative technologies that have influenced scholarly communication and research assessment. Advances in ML, NLP, and LLMs have significantly changed discovery of literature, generation of manuscripts, analysis of data, and dissemination of findings. AI-driven applications are integrated into academic publishing workflows, and offers considerable improvements in efficacy and decision making. But this rise also brings potential pitfalls in terms of fairness, reliability and governance. So, the use of AI, makes it an enabler or a disruptor of trustworthy and fair research assessment (Bhattacharya, 2026; van Dis *et al*., 2023).

Positive Roles of AI in Research Assessment

AI can improve literature screening. The expansion of scientific publications has rendered manual evidence synthesis difficult. This is evident in disciplines that produce innumerable articles yearly. AI-assisted systematic review tools can identify relevant publications, classify research themes, remove duplicate records, and prioritize articles for assessment by reviewer automatically. By improving level of consistency and reproducibility, these capabilities reduce the time for literature reviews. Although confirmation by experts is essential, AI enables researchers to manage large scale evidence more efficiently compared to manual methods (van Dis *et al*., 2023).

AI can identify fake manuscripts within scholarly publishing. ML algorithms analyze similarities in texts, image manipulation, statistical inconsistencies, and unusual publication patterns. The AI scanning systems can pick up plagiarized articles, double

published images, modified data and suspicious peer review processes. The power to discover research malpractice at its inception boosts editorial quality control procedures (Notman, 2025).

Another rapidly evolving use of AI is identifying paper mills. A paper mill is an organization that produces fraudulent scientific research papers, thereby devaluing research. AI can assist in detecting distinctive linguistic styles, redundant author or institution names, manipulated images, aberrant citation patterns, synchronized publishing behavior. Publishers use AI surveillance technologies to protect research from paper mills and keep records of scientific activity pristine (Notman, 2025).

AI also excels in spotting abnormal citations. AI-based models using citation network analysis can effectively pinpoint unusual citation patterns such as hyper-self-citation, citation cartels, citation trading, citation rings, etc. Using huge volumes of bibliometric data, AI can easily differentiate typical citation behavior from inappropriate methods adopted by certain researchers to boost citation impact and inflate bibliometrics. Such methods enhance reliability of bibliometrics and help promote responsible research assessment principles (Bhattacharya, 2026).

Negative Roles of AI in Research Assessment

Despite the benefits mentioned above, AI raises several major issues for research assessment. Among the most frequently reported issue is generation of fake references. LLMs produce citations that look legitimate but correspond to fake articles, wrong authors, fictitious titles, or wrong publication details. Researchers failing to verify these references risk introducing false information into scholarly literature. This contaminates bibliographic databases and compromises ensuing bibliometric analyses (Walters and Wilder, 2023).

This challenge is further intensified by AI-generated manuscripts. Generative AI systems can produce an entire research paper or a part thereof in minutes. While these tools help improve efficiency of writing and quality of language, they also produce low quality or inadequately verified manuscripts. Without suitable human oversight, content generated by AI runs the risk of factual inaccuracies, logical inconsistencies, or unsupported interpretations that undermine scientific credibility (van Dis *et al*., 2023).

Hallucinated citations are another major issue of AI in assessing research. Here, AI systems generate non-existent or incorrect bibliographic information. Walters and Wilder (2023) found that conversational AI models provide fake citations while responding to academic queries. Failure to undertake manual verification results in non-detection of fake references. These reduce the trustworthiness of literature reviews, systematic reviews, and citation databases.

The ability of AI to rapidly generate manuscripts identical to legitimate scientific papers also indirectly contributes to the expansion of paper mills. Technology like automated writing systems and image and citation generators has reduced the costs of creating fake articles. This has increased editors' workloads, challenged peer-review processes and undermined the accuracy of bibliometric measurements. This requires strengthening of AI assisted quality assurance mechanisms (Bhattacharya, 2026; Notman, 2025).

ETHICAL CONCERNS IN AI-ASSISTED RESEARCH ASSESSMENT

The integration of AI into research assessment raises several ethical concerns that relate to transparency, accountability, and verification.

Transparency requires authors, reviewers, and publishers to give a declaration stating the extent to which AI tools have been used in generating the scientific output. Such declarations help readers and assessors to differentiate between human and AI aided outputs. This helps maintain confidence in scholarly communication (COPE Council, 2024). Further, while the predictive capabilities of LLMs may lead to supplementing traditional citation-based indicators, the inherent complexity and lack of openness in decision making processes require continuous reliance on human expert review (Thelwall, 2026).

AI systems are responsible for the accuracy, originality, or ethical implications of research outputs. This increases the importance of accountability. Accountability requires establishing frameworks that addresses biases, errors, or ethical breaches that are intrinsic to algorithmic decision making (Kharipova *et al*., 2024). Authors are required to guarantee the validity of contents generated by AI. Further, they must ensure that content has been adequately referenced and does not involve plagiarism or fabricated content. Similarly, editors and institutions must define acceptable usages of AI during all stages of research (UNESCO, 2022). As technologies advance, interdisciplinary cooperation of researchers, ethicists and policy makers will become increasingly important to formulate adequate flexible policies to uphold scholarly integrity (Saeidnia *et al*., 2024).

Verification helps maintain integrity of research assessment. All scientific contents generated by AI must face human verification prior to publication. The function of AI should not be as an autonomous assessor but as a decision support tool. Combining AI enabled efficacy with expert oversight can maximize technological innovation while safeguarding integrity of research, bibliometric reliability, and public confidence in systems of scientific assessment (Bhattacharya, 2026; van Dis *et al*., 2023).

RESPONSIBLE BIBLIOMETRICS

The growing concerns regarding the misuse of research metrics have led to developing responsible bibliometrics. This approach advocates ethical, transparent, and context sensitive use of quantitative indicators in assessing research. Responsible bibliometrics does not reject bibliometric measures; rather it promotes their proper application along expert judgment to ensure fair and inclusive assessment of research quality (Bhattacharya, 2026).

The Leiden Manifesto, proposed by Hicks *et al.* (2015), is among the most significant frameworks. This manifesto outlines ten principles for the responsible use of research metrics. The principles include (i) quantitative assessment should support qualitative expert assessment, (ii) indicators should reflect research missions, (iii) assessment of research within disciplinary contexts, (iv) data collection and analytical methods should be kept open, transparent, and simple, (v) researchers should be allowed to verify assessment data, (vi) accounting for inter-field variations, (vii) researchers should be assessed based on holistic, qualitative review, (viii) avoidance of false precision, (ix) recognition of the systemic effects of assessment, and (x) regular review of indicators to prevent unintended consequences. Collectively, these principles encourage assessors to construe metrics carefully and avoid crude numerical comparisons.

DORA, introduced in 2013, is another important initiative. DORA discourages using JIF as a proxy for assessing the quality of individual researchers or articles. It recommends assessing research on its inherent scientific merit while considering diverse outputs like datasets, software, policy contributions, and common impact. DORA encourages institutions and funding agencies to adopt an open and fair assessment criterion that recognizes the diversity of scholarly contributions (DORA, 2013).

More recently, CoARA has expanded efforts aimed at reforming research assessment. Established in 2022, CoARA advocates replacing linear metric-based assessment systems with multidimensional systems, encompassing qualitatively based peer review, careful and responsible use of quantitative indicators, open science processes, and acknowledgement of varied scholarly works. CoARA has engaged a range of academic institutions, funding institutions, and research organizations in efforts to promote sound assessment practices (CoARA, 2022).

In similar vein, the Metric Tide report introduced five fundamental principles of responsible metrics. These include robustness, ensuring the use of high-quality and reliable data, requiring open and verifiable assessment methods, affirming that metrics should support expert judgment, encouraging multiple indicators to reflect disciplinary and institutional differences, and promoting continuous assessment of the effects and unintended consequences of assessment systems (Wilsdon *et al.*, 2015).

These models illustrate that one single bibliometric indicator cannot measure the quality of research. The quantitative evidence can be strong evidence of a researcher's academic impact, but fails to encompass aspects of newness, robust methodology, trans-disciplinarity, impact on society and ethical practices. Effective assessment of research, therefore, involves a mix of bibliometric metrics, supported by robust peer review, and considering the context in which the research has taken place to promote fair, transparent, and balanced assessment systems, while ensuring research integrity (Bhattacharya, 2026; Rushforth and Hammarfelt, 2023).

FUTURE DIRECTIONS

The future of research assessment requires developing transparent, inclusive, ethically grounded assessment systems that can adapt to evolving scholarly landscape. With bibliometric indicators, emerging developments like narrative bibliometrics, Open Science, AI governance, transparent indicators, context sensitive assessment, and human AI association are expected to redefine measurement of quality and impact of research. These approaches, besides strengthening research integrity and public trust, also seek to address the limitations of conventional metric-based assessment systems (Bhattacharya, 2026).

Narrative Bibliometrics represents one such encouraging development that attempts to integrate quantitative indicators and qualitative narratives into the researcher's profile. This allows for an enriched representation of their contributions where instead of merely citing statistics, the importance, interdisciplinary aspects, societal relevance, and actual impacts of their works are explained based on relevant bibliometric data. Narrative bibliography tries to reconcile the research assessment process with other academic activities of the researcher. This helps curbs the excessive reliance on numerical indicators (Torres-Salinas *et al.*, 2024).

The increasing adoption of Open Science principles is likely to influence future directions in assessing research. By promoting openness, Open science initiatives aim to improve the accessibility and verifiability of scientific knowledge. Future research assessment systems will recognize contributions to science beyond traditional publications, like sharing of datasets, software, protocols, and other research outputs. This encourages interdisciplinary and transnational research collaborations. It aligns with UNESCO's call for more comprehensive and equitable research assessment frameworks (UNESCO, 2022).

To ensure responsible AI-driven research assessment, all stakeholders must put forward adequate AI governance to safeguard the research quality. Clear policies on ethical application of AI technologies in all stages of academic writing process are required. It means that governance framework must imply the necessity of disclosing the usage of AI tools by researchers during the whole workflow from manuscript preparation to publication. Additionally, these guidelines may cover issues like responsibility

assignment between authors and reviewers regarding AI-generated content, ways for AI-driven misinformation detection, fabrication of citations, and manipulation of research output, etc. Such measures are needed to maintain credibility and integrity in scientific evaluation amidst rapid adoption of artificial intelligence (COPE Council, 2024; van Dis *et al*., 2023).

Future research assessments will place an emphasis on transparent indicators. These should be supported by reliable, verifiable, and openly accessible data. Methodologies for calculating bibliometric indicators need to be documented, reproducible, and routinely assessed for potential bias. Transparent indicators allow researchers to comprehend how assessment outcomes can be reached and create opportunities to find and rectify inaccuracies. This ensures greater confidence in assessment processes and encourages appropriate usage of research metrics (Hicks *et al*., 2015; Wilsdon *et al*., 2015).

It is impossible to assess the quality of research using the same standards across different academic fields and institutional types. Citation behavior and other research output indicators differ across scientific disciplines. There are cases where universal application of certain assessment measures is inappropriate or useless. So, future frameworks need to consider how factors like field-specific norms, institutional mission statements, common impacts, career trajectory differences, and geographical considerations influence the interpretation of quantitative indicators. This would allow researchers to better balance both quantitative and subjective qualitative assessments through peer review (Aksnes *et al*., 2019; Bhattacharya, 2026).

Effective human-AI collaboration is another viable alternative. With AI supporting literature screening, citation analysis, detection of fraud, and data processing, human experts can provide ethical judgment, and disciplinary expertise that cannot be replicated by machines. A collaborative model, with AI functioning as a support tool under human supervision, offers great potential to improve efficiency while maintaining accountability, openness, and integrity. Incorporation of technological innovation with responsible governance and expert assessment will help future assessment systems recognize the diverse contributions of researchers and encourage responsible and inclusive global research ecosystem (Bhattacharya, 2026; Rushforth and Hammarfelt, 2023).

RECOMMENDATIONS

Responsible research assessment requires harmonized efforts from all stakeholders.

Researchers should avoid immoral metric driven practices and focus on quality of research, integrity, openness, and responsible use of AI.

Universities should use assessment systems that combine bibliometric indicators with peer review and acknowledge various types of scholarly outputs.

Funding agencies should reduce dependence on single metrics and use field sensitive, open assessment criteria that value common impact and open science practices.

Publishers should implement AI assisted integrity checks, detect fraudulent publications, and ensure disclosure of AI use.

Governments should form policies that promote responsible bibliometrics, invest in research integrity initiatives, encourage Open Science, and support standardized yet context sensitive assessment frameworks.

Collectively, these measures will create equitable, transparent, and reliable research assessment systems.

CONCLUSION

Despite finding wide usage in assessing research, bibliometric indicators are insufficient as standalone measures of scientific quality. Indiscriminate use can change research behaviors that undermine academic values and result in biased interpretations of ethical practices. AI provides exciting prospects yet poses several challenges. Future approaches should combine quantitative indicators with expert peer review, qualitative evidence, and contextual interpretation. Adherence to these principles will promote research integrity and foster confidence. Balanced human-AI collaboration is the best alternative towards credible, inclusive, and sustainable research assessment.

ABBREVIATIONS

AI: Artificial Intelligence; **CoARA:** Coalition for Advancing Research Assessment; **DORA:** San Francisco Declaration on Research Assessment; **FWCI:** Field-Weighted Citation Impact; **JIF:** Journal Impact Factor; **LLM:** Large Language Model; **ML:** Machine Learning; **NLP:** Natural Language Processing; **SJR:** SCImago Journal Rank; **SNIP:** Source Normalized Impact per Paper; **WoS:** Web of Science.

CONFLICT OF INTEREST

The author declare that there is no conflict of interest.

REFERENCES

Aksnes, D. W., Langfeldt, L., & Wouters, P. (2019). Citations, citation indicators, and research quality: An overview of basic concepts and theories. Sage Open, 9(1), 2158244019829575. https://doi.org/10.1177/2158244019829575

Bhattacharya, S. (2026). Bibliometrics at the crossroads: When the measure becomes the target. Journal of Scientometric Research, 15(1), 1. https://doi.org/10.5530/jscires.20260001

Belter, C. W. (2015). Bibliometric indicators: Opportunities and limits. Journal of the Medical Library Association: JMLA, 103(4), 219-221. https://doi.org/10.3163/1536-5050.103.4.014

Biagioli, M., & Lippman, A. (Eds.). (2020). *Gaming the metrics: Misconduct and manipulation in academic research*. The MIT Press. https://doi.org/10.7551/mitpress/11087.001.0001

CoARA. (2022). Reforming research assessment: The Agreement is now final. https://research-and-innovation.ec.europa.eu/news/all-research-and-innovation-news/reforming-research-assessment-agreement-now-final-2022-07-20_en

COPE Council. (2024). Authorship and AI tools. Committee on Publication Ethics. https://doi.org/10.24318/cCVRZBms

Daraio, C., Glänzel, W., & Gorraiz, J. (2025). A responsible framework for an appropriate bibliometric-based research assessment. In *Proceedings of the 20th International Conference on Scientometrics & Informetrics*. https://doi.org/10.51408/issi2025_160

DORA. (2013). San Francisco declaration on research assessment. Responsible Research Assessment Documents. https://sfdora.org/read/

Elsevier. (2026). *Field-Weighted Citation Impact (FWCI) metrics*. Elsevier Pure Help Center. https://helpcenter.pure.elsevier.com/field-weighted-citation-impact-fwci-metrics

Garfield, E. (1972). Citation Analysis as a Tool in Journal Evaluation: Journals can be ranked by frequency and impact of citations for science policy studies. Science, 178(4060), 471-479. https://doi.org/10.1126/science.178.4060.471

Hicks, D., Wouters, P., Waltman, L., De Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden manifesto for research metrics. Nature, 520(7548), 429-431. https://doi.org/10.1038/520429a

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. Proceedings of the National Academy of Sciences, 102(46), 16569-16572. https://doi.org/10.1073/pnas.0507655102

Kar, S. (2026). Rethinking research evaluation in India: Is the ecosystem ready for a responsible reform agenda? Journal of Scientometric Research, 14(3), 925-928. https://doi.org/10.5530/jscires.20250068

Kharipova, R., Khaydarov, I., Akramova, S., Lutfullaeva, D., Saidov, S., *et al.* (2024). The role of artificial intelligence technologies in evaluating the veracity of scientific research. Journal of Internet Services and Information Security, 14(4), 554-568. https://doi.org/10.58346/JISIS.2024.I4.035

Notman, N. (2025, May 14). AI tools combat paper mill fraud in scientific publishing as peer review system struggles. *Chemistry World*. https://www.chemistryworld.com/features/ai-tools-tackle-paper-mill-fraud-overwhelming-peer-review/4022253.article

Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (2010). *Altmetrics: A manifesto*. DigitalCommons@University of Nebraska - Lincoln. https://digitalcommons.unl.edu/scholcom/185

Rushforth, A., & Hammarfelt, B. (2022). The rise of 'responsible metrics' as a professional reform movement: A collective action frames account, 4(4), 879-897. https://doi.org/10.31235/osf.io/cdmqz

Saeidnia, H. R., Hosseini, E., Abdoli, S., & Ausloos, M. (2025). Unleashing the power of AI: A systematic review of cutting-edge techniques in AI-enhanced scientometrics, webometrics and bibliometrics. Library Hi Tech, 43(4-5), 1292-1314. https://doi.org/10.1108/LHT-10-2023-0514

Thelwall, M. (2026). Large language models and responsible research evaluation: An extension of the Leiden Manifesto. *Scientometrics*. Advance online publication. https://doi.org/10.1007/s11192-026-05552-x

Torres-Salinas, D., Orduña-Malea, E., Delgado-Vázquez, Á., Gorraiz, J., & Arroyo-Machado, W. (2024). Foundations of narrative bibliometrics. Journal of Informetrics, 18(3), 101546. https://doi.org/10.1016/j.joi.2024.101546

UNESCO. (2022). Recommendation on the Ethics of Artificial Intelligence. https://unesdoc.unesco.org/ark:/48223/pf0000381137

Van Dis, E. A. M., Bollen, J., Zuidema, W., Van Rooij, R., & Bockting, C. L. (2023). ChatGPT: Five priorities for research. Nature, 614(7947), 224-226. https://doi.org/10.1038/d41586-023-00288-7

Walters, W. H., & Wilder, E. I. (2023). Fabrication and errors in the bibliographic citations generated by ChatGPT. Scientific Reports, 13(1), 14045. https://doi.org/10.1038/s41598-023-41032-5

Wilsdon, J. (2015). *The metric tide: Independent review of the role of metrics in research assessment and management*. SAGE Publications. https://doi.org/10.4135/9781473978782

Cite this article: Maurya A. Responsible Research Assessment in the Age of Artificial Intelligence: Challenges and Future Directions. Info Res Com. 2027;4(1):69-76.