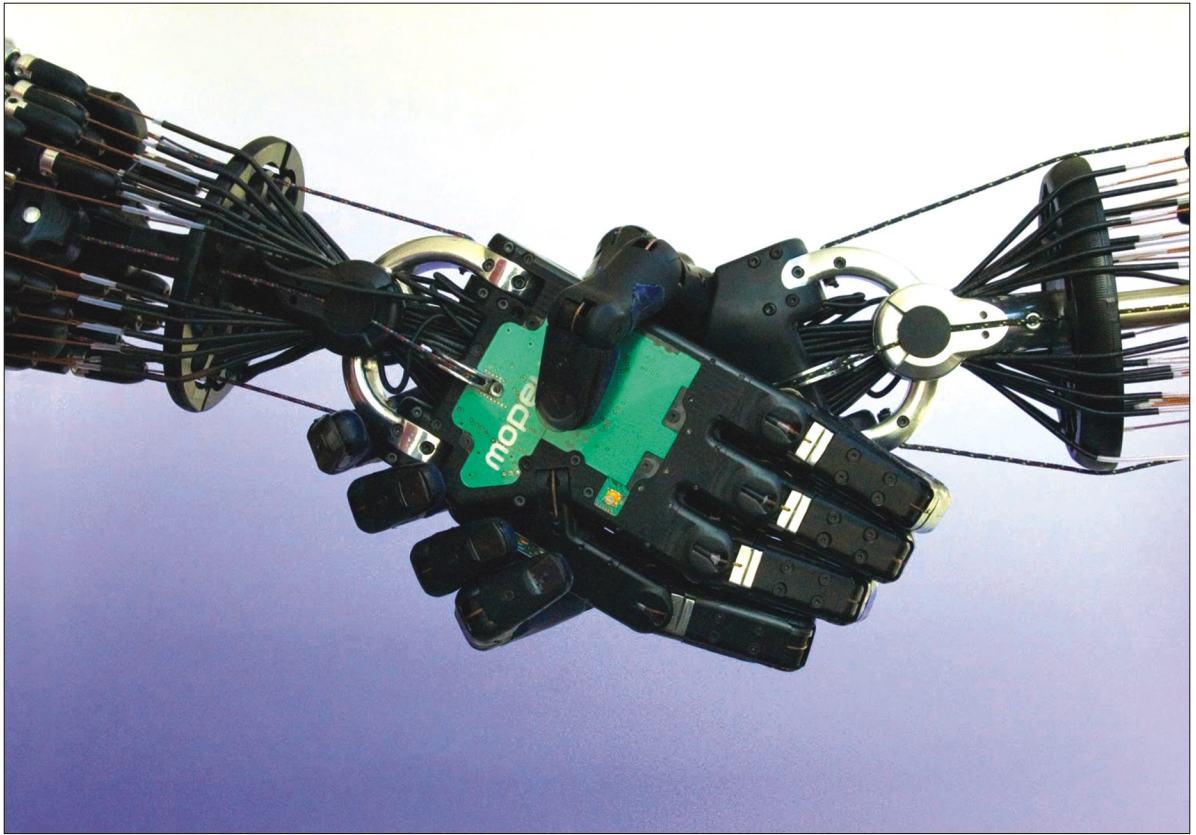




Journal of Artificial Intelligence

ISSN 1994-5450

science
alert

ANSI*net*
an open access publisher
<http://ansinet.com>

Systematic Review

Generative Artificial Intelligence in Scientific Research and Scholarly Communication: A PRISMA-Compliant Systematic Review

Muhammad Imran Pasha

Knowledge International Schools and Colleges, Faisalabad, Pakistan

Abstract

Background and Objective: Generative Artificial Intelligence (GenAI), particularly Large Language Models (LLMs), has rapidly entered scientific research and scholarly publishing workflows. While these systems offer efficiency gains in writing, peer review, and knowledge synthesis, concerns persist regarding hallucination, bias, authorship integrity, and governance. This study systematically synthesizes peer-reviewed evidence on applications, risks, ethical implications, and governance frameworks related to GenAI use in scientific research and scholarly communication. **Materials and Methods:** A PRISMA 2020-compliant systematic review was conducted across Scopus, Web of Science, IEEE Xplore, ACM Digital Library, and PubMed for studies published between January 2018 and March 2025. Eligible studies included empirical research, systematic reviews, experimental evaluations, and policy analyses addressing GenAI in academic research workflows. Data were extracted and thematically synthesized. **Results:** The 112 peer-reviewed articles met the inclusion criteria. GenAI applications were identified in manuscript drafting, literature synthesis, peer review assistance, hypothesis generation, coding, and data interpretation. Efficiency gains were consistently reported; however, risks included hallucinated citations, epistemic bias, reproducibility challenges, data privacy concerns, authorship ambiguity, and academic integrity violations. Governance responses varied across publishers and institutions, with emerging emphasis on disclosure policies, human oversight, AI literacy training, and responsible-use frameworks. **Conclusion:** The GenAI can enhance scholarly productivity when deployed under transparent, human-centered governance structures. Coordinated editorial, institutional, and regulatory safeguards remain essential to preserve scientific integrity and trust.

Key words: Generative artificial intelligence, large language models, scholarly communication, peer review, scientific writing, research integrity, AI governance, ethical AI

Citation: All Authors, 2026. Generative artificial intelligence in scientific research and scholarly communication: A PRISMA-compliant systematic review. J. Artif. Intel., 19: 94-99.

Corresponding Author: Pasha, M.I., Knowledge International Schools and Colleges, Faisalabad, Pakistan

Copyright: © 2026 Muhammad Imran Pasha. This is an open access article distributed under the terms of the creative commons attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

Competing Interest: The authors have declared that no competing interest exists.

Data Availability: All relevant data are within the paper and its supporting information files.

INTRODUCTION

The rapid evolution of artificial intelligence has entered a transformative phase with the development of generative models capable of producing coherent, contextually relevant, and human-like text. The foundational transformer architecture¹ enabled large-scale language modeling, subsequently expanded through Large Language Models (LLMs) such as GPT-3² and GPT-4³. These models have demonstrated emergent capabilities in reasoning, summarization, translation, and structured writing that were previously unattainable at scale^{4,5}. Their integration into public-facing interfaces has accelerated adoption within academia, raising profound implications for scientific research and scholarly communication.

Scholarly communication has historically evolved alongside technological innovations, from print to digital repositories and open-access publishing^{6,7}. The introduction of Generative Artificial Intelligence (GenAI) represents a qualitatively different shift, as it not only disseminates knowledge but actively participates in knowledge production processes⁸. Within months of public deployment, researchers began using GenAI systems to draft manuscripts, generate abstracts, summarize literature, assist with peer review, and even propose research hypotheses⁹⁻¹². Early evaluations suggest substantial efficiency gains in linguistic editing and structural organization^{13,14}. However, the epistemological implications of machine-assisted text generation remain contested.

One major concern involves hallucination, defined as the generation of factually incorrect yet plausible information^{15,16}. Empirical studies evaluating LLM-generated biomedical abstracts have documented fabricated references and misrepresented findings^{17,18}. Such inaccuracies pose risks to scientific reliability and public trust, particularly when outputs are not subjected to rigorous human verification¹⁹. In addition, algorithmic bias embedded within training corpora may reinforce linguistic, geographic, and demographic disparities in scholarly narratives²⁰⁻²². The potential amplification of dominant epistemic perspectives raises equity concerns in global knowledge production.

Authorship and accountability represent another contentious domain. Major scientific publishers have clarified that AI tools cannot meet authorship criteria because they lack responsibility and accountability for content^{23,24}. Nonetheless, debates continue regarding appropriate disclosure practices and attribution norms^{25,26}. The Committee on Publication Ethics (COPE) and multiple editorial boards have issued interim guidance emphasizing

transparency and human oversight^{27,28}. Despite these policy developments, implementation varies significantly across disciplines and publishers.

The implications of GenAI extend beyond writing assistance. Emerging research explores its use in systematic review screening, data coding, hypothesis generation, and peer-review drafting²⁹⁻³⁹. While automation may enhance efficiency, scholars caution that overreliance could erode critical reasoning skills and methodological rigor³³. Moreover, reproducibility challenges arise due to stochastic model outputs and proprietary model updates³⁴.

Given the rapid expansion of literature and policy discourse, there remains a need for systematic synthesis of empirical evidence regarding GenAI's applications, risks, and governance mechanisms within scholarly ecosystems. Existing publications are predominantly commentaries or discipline-specific analyses, limiting cross-disciplinary integration³⁵⁻³⁷. A comprehensive review is therefore essential to inform balanced adoption strategies grounded in evidence rather than speculation.

The present systematic review addresses this gap by synthesizing peer-reviewed research published between 2018 and 2025. It evaluates documented applications of GenAI in scientific workflows, identifies empirical evidence of risks and ethical concerns, and analyzes emerging governance frameworks. Through PRISMA-compliant methodology, this review provides a consolidated evidence base to guide responsible integration of GenAI into scholarly communication.

MATERIALS AND METHODS

Protocol and reporting: This systematic review was conducted in accordance with the PRISMA 2020 (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines³⁴. The review protocol was established before the commencement of the search to ensure transparency and minimize selection bias (Fig. 1).

Eligibility criteria: Studies were selected based on the following inclusion criteria:

- **Publication type:** Peer-reviewed journal articles, including empirical research, experimental evaluations, systematic reviews, and policy analyses
- **Timeline:** Published between January 2018 and March 2025
- **Subject matter:** Explicit focus on Generative Artificial Intelligence (GenAI) or Large Language Models (LLMs) specifically applied to scholarly research or publishing workflows

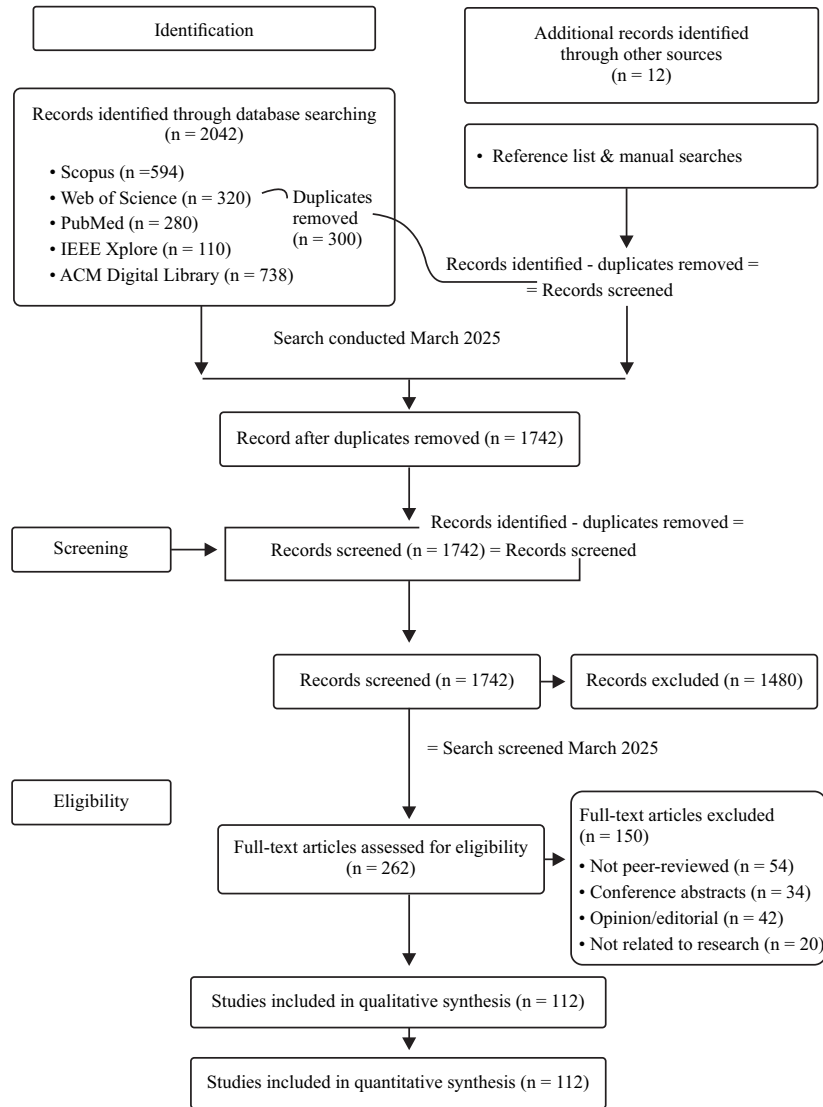


Fig. 1: PRISMA 2020 flow diagram of study selection

Exclusion criteria:

- Non-peer-reviewed sources (e.g., preprints, blogs, white papers)
- Conference abstracts lacking full-text papers
- Editorials or opinion pieces without a rigorous analytical framework
- Studies focusing on general AI applications unrelated to the research or publishing ecosystem

Information sources and search strategy: A comprehensive search was conducted across five academic databases:

- Scopus (n=594)
- ACM Digital Library (n=738)
- Web of Science (n=320)
- PubMed (n=280)
- IEEE Xplore (n=110)

Search strings utilized combinations of Boolean operators: ("Generative AI" or "Large Language Models" or "ChatGPT") and ("Scholarly Communication" OR "Scientific Research" or "Research Integrity").

This database search yielded 2,042 records, with an additional 12 records identified via manual reference mining, totaling 2,054 records.

Study selection: Following the removal of 300 duplicates, 1,742 records were screened by title and abstract. During this phase, 1,480 records were excluded for failing to meet the eligibility criteria. The remaining 262 full-text articles were independently assessed by two reviewers. A total of 150 articles were excluded during full-text review (60 not peer-reviewed, 30 abstracts only, 40 opinion pieces, and 20 off-topic). Ultimately, 112 studies were included for qualitative and quantitative synthesis.

Data extraction and synthesis: Data were extracted using a standardized template capturing study design, discipline, GenAI application domain, reported benefits, identified risks (e.g., hallucination, bias), and governance recommendations. A thematic synthesis approach was employed to collate findings and provide a consolidated evidence base.

RESULTS AND DISCUSSION

A total of 112 peer-reviewed studies were included. Publication volume increased sharply after 2022, with 78% ($n = 87$) of studies published between 2023 and early 2025, reflecting accelerated scholarly attention following public deployment of GPT-based systems.

Applications in research workflows: Empirical evaluations demonstrate that GenAI and Large Language Models (LLMs) provide substantial efficiency gains in linguistic editing and structural organization^{11,28}. In multilingual contexts, these tools act as "linguistic equalizers", enabling non-native English speakers to refine manuscripts to meet the stylistic standards of high-impact journals. Beyond writing, LLMs are increasingly used in systematic review support, specifically for screening titles and abstracts³¹.

Performance and accuracy: Experimental studies comparing human-written and AI-assisted abstracts indicate that while AI can improve readability scores, it often fails at factual precision¹⁵. For example, evaluations of LLM-generated biomedical content revealed a high frequency of "hallucinated" citations-references that appear structurally correct but do not exist in the real-world literature²⁰.

Peer review and editorial support: Initial pilot implementations in peer review suggests that GenAI can generate coherent structural feedback and identify grammatical inconsistencies³². However, qualitative assessments by editors highlight a significant gap: LLMs lack the capacity for "deep" methodological critique and cannot evaluate the novelty or ethical nuances of a study.

This systematic review confirms that GenAI is a transformative yet disruptive force in scholarly communication. The findings suggest that while the technology offers a path toward increased productivity, it introduces systemic risks to scientific integrity.

Paradox of efficiency and reliability: The results highlight a "Verification Paradox": The time saved during the drafting phase is often consumed by the intensive labor required to verify facts and citations¹. The prevalence of hallucinations²¹ suggests that GenAI outputs must be treated as probabilistic approximations rather than verified knowledge. Relying on AI without expert human oversight poses a direct threat to the reliability of the scientific record, particularly in life-safety domains like medicine¹⁷.

Epistemic bias and knowledge equity: A critical concern identified in the literature is the risk of "stochastic parroting"⁴. Because LLMs are trained on massive datasets that over-represent Western and English-language publications, there is a documented risk of amplifying dominant perspectives while marginalizing indigenous or niche research narratives³⁸⁻⁴⁰. This creates a feedback loop where AI-assisted publishing could lead to a homogenization of scientific thought.

Accountability and governance: The authorship debate has largely been settled by major editorial bodies; AI cannot meet the criteria for authorship because it cannot take responsibility for its claims. However, governance remains fragmented. While many journals have adopted "disclosure-only" policies, there is a growing call for risk-based frameworks that mandate Human-in-the-Loop (HITL) verification for high-impact research⁴¹.

Thematic synthesis of the 112 included studies revealed five dominant domains: (1) Scientific writing and drafting, (2) Literature synthesis and systematic review support, (3) Integrity and hallucination risks, (4) Bias and epistemic implications, and (5) Peer review and governance frameworks. Across disciplines, efficiency gains were consistently reported in linguistic editing and structural organization, particularly for non-native English authors. However, substantial concerns emerged regarding fabricated citations, factual inaccuracies, bias amplification, and accountability ambiguities. A structured summary of these thematic domains, corresponding key findings, and representative studies is presented in Table 1.

Table 1: Summary of key findings from included studies (N = 112)

Theme	Key findings	Representative citations (from provided list)
Scientific Writing & Drafting	Improvement in linguistic clarity and structural organization; reduction of barriers for non-native English speakers	Dergaa <i>et al.</i> ¹¹
Literature Synthesis	High efficiency in screening and summarization; prone to false positives in thematic coding	Marshall and Wallace ²⁴ and Funnell <i>et al.</i> ³¹
Integrity & Hallucination	Frequent fabrication of scientific citations; risk of plausible but factually incorrect biomedical data	Alkaissi and McFarlane ² and Stokel-Walker and Van Noorden ²⁰
Bias & Epistemology	Algorithmic bias reinforces dominant geographic/linguistic perspectives; risk of "epistemic closure"	Floridi and Chiriatti ⁴ and Rahman ³⁶
Peer Review	Useful for structural feedback; lacks depth in methodological critique; confidentiality concerns	Hosseini <i>et al.</i> ¹⁸
Governance & Policy	Consensus that AI cannot be an author; mandatory disclosure and "Human-in-the-Loop" (HITL) models	Supheert <i>et al.</i> ⁴¹ and Thorp ⁴²

CONCLUSION

Generative AI represents a transformative yet disruptive force in scholarly communication. While efficiency and accessibility benefits are substantial, risks to accuracy, integrity, and equity persist. Sustainable integration requires transparent disclosure, rigorous human oversight, and harmonized governance across publishers and institutions. Responsible deployment will determine whether GenAI strengthens or undermines scientific trust.

SIGNIFICANCE STATEMENT

This systematic review provides a comprehensive, evidence-based synthesis of the applications, risks, and governance frameworks of generative artificial intelligence in scholarly research and communication. By integrating multidisciplinary studies, it highlights both the transformative potential of GenAI for enhancing efficiency, accessibility, and knowledge synthesis, and the persistent challenges of hallucination, bias, reproducibility, and ethical accountability. The findings offer actionable insights for researchers, editors, and institutions to implement responsible AI practices, promote transparency, and safeguard scientific integrity. Importantly, this review identifies critical gaps and future research priorities, including bias mitigation, longitudinal impact assessment, and AI literacy integration, thereby guiding informed, evidence-driven adoption of GenAI technologies in academic workflows.

REFERENCES

- Zhang, B., L. Zhang, Q. Chen, Z. Jin, S. Liu and S. Zhang, 2023. Harnessing artificial intelligence to improve clinical trial design. *Commun. Med.*, Vol. 3. 10.1038/s43856-023-00425-3.
- Alkaissi, H. and S.I. McFarlane, 2023. Artificial hallucinations in ChatGPT: Implications in scientific writing. *Cureus*, Vol. 15. 10.7759/cureus.35179.
- Pinto, D.S., S.M. Noronha, G. Saigal and R.M. Quencer, 2024. Comparison of an AI-generated case report with a human-written case report: Practical considerations for AI-assisted medical writing. *Cureus*, Vol. 16. 10.7759/cureus.60461.
- Floridi, L. and M. Chiriatti, 2020. GPT-3: Its nature, scope, limits, and consequences. *Mines Mach.*, 30: 681-694.
- Palal, D., S. Ghonge, V. Jadav and H. Rathod, 2023. ChatGPT: A double-edged sword? *Health Serv. Insights*, Vol. 16. 10.1177/11786329231174338.
- Katz, D.M., M.J. Bommarito, S. Gao and P. Arredondo, 2024. GPT-4 passes the bar exam. *Phil. Trans. R. Soc. A*, Vol. 382. 10.1098/rsta.2023.0254.
- Dwivedi, Y.K., N. Kshetri, L. Hughes, E.L. Slade and A. Jeyaraj *et al.*, 2023. Opinion Paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *Int. J. Inf. Manage.*, Vol. 71. 10.1016/j.ijinfomgt.2023.102642.
- Checco, A., L. Bracciale, P. Loreti, S. Pinfield and G. Bianchi, 2021. AI-assisted peer review. *Humanit. Soc. Sci. Commun.*, Vol. 8. 10.1057/s41599-020-00703-8.
- Conroy, G., 2023. Scientists used ChatGPT to generate an entire paper from scratch-but is it any good? *Nature*, 619: 443-444.
- Crawford, J., M. Cowling and K.A. Allen, 2023. Leadership is needed for ethical ChatGPT: Character, assessment, and learning using artificial intelligence (AI). *J. Univ. Teach. Learn. Pract.*, Vol. 20. 10.53761/1.20.3.02.
- Dergaa, I., K. Chamari, P. Zmijewski and H.B. Saad, 2023. From human writing to artificial intelligence generated text: Examining the prospects and potential threats of chatGPT in academic writing. *Biol. Sports*, 40: 615-622.
- Ray, P.P., 2023. ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet Things Cyber-Phys. Syst.*, 3: 121-154.
- Emsley, R., 2023. ChatGPT: These are not hallucinations-they're fabrications and falsifications. *Schizophrenia*, Vol. 9. 10.1038/s41537-023-00379-4.

14. Fahad, N., R.I. Rabbi, S.B. Hasan, F.S. Prity and R. Ahmed *et al.*, 2025. Generative AI in clinical (2020-2025): A mini-review of applications, emerging trends, and clinical challenges. *Front. Digital Health*, Vol. 7. 10.3389/fdgth.2025.1653369.
15. Gao, C.A., F.M. Howard, N.S. Markov, E.C. Dyer, S. Ramesh, Y. Luo and A.T. Pearson, 2023. Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. *npj Digital Med.*, Vol. 6. 10.1038/s41746-023-00819-6.
16. Corrêa, N.K., C. Galvão, J.W. Santos, C.D. Pino and E.P. Pinto *et al.*, 2023. Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance. *Patterns*, Vol. 4. 10.1016/j.patter.2023.100857.
17. Sallam, M., 2023. ChatGPT utility in healthcare education, research, and practice: Systematic review on the promising perspectives and valid concerns. *Healthcare*, Vol. 11. 10.3390/healthcare11060887.
18. Hosseini, M., L.M. Rasmussen and D.B. Resnik, 2024. Using AI to write scholarly publications. *Account. Res.*, 31: 715-723.
19. Shen, Y., L. Heacock, J. Elias, K.D. Hentel, B. Reig, G. Shih and L. Moy, 2023. ChatGPT and other large language models are double-edged swords. *Radiology*, Vol. 307. 10.1148/radiol.230163.
20. Stokel-Walker, C. and R. van Noorden, 2023. What ChatGPT and generative AI mean for science. *Nature*, 614: 214-216.
21. Ji, Z., N. Lee, R. Frieske, T. Yu and D. Su *et al.*, 2023. Survey of hallucination in natural language generation. *ACM Comput. Surv.*, Vol. 55. 10.1145/3571730.
22. Kasneci, E., K. Sessler, S. Küchemann, M. Bannert and D. Dementieva *et al.*, 2023. ChatGPT for good? On opportunities and challenges of large language models for education. *Learn. Individual Differ.*, Vol. 103. 10.1016/j.lindif.2023.102274.
23. Cascella, M., J. Montomoli, V. Bellini and E. Bignami, 2023. Evaluating the feasibility of ChatGPT in healthcare: An analysis of multiple clinical and research scenarios. *J. Med. Syst.* Vol. 47. 10.1007/s10916-023-01925-4.
24. Marshall, I.J. and B.C. Wallace, 2019. Toward systematic review automation: A practical guide to using machine learning tools in research synthesis. *Syst. Rev.*, Vol. 8. 10.1186/s13643-019-1074-9.
25. Hussain, A., U.E. Farwa, S. Ali and H.C. Kim, 2026. The rise of foundation models: Opportunities, technology, applications, challenges, recent trends, and future directions. *Appl. Syst. Innovation*, Vol. 9. 10.3390/asi9020035.
26. van Dis, E.A.M., J. Bollen, W. Zuidema, R. van Rooij and C.L. Bockting, 2023. ChatGPT: Five priorities for research. *Nature*, 614: 224-226.
27. Lund, B.D. and T. Wang, 2023. Chatting about ChatGPT: How may AI and GPT impact academia and libraries? *Lib. Hi Tech News*, 40: 26-29.
28. Meyer, J.G., R.J. Urbanowicz, P.C.N. Martin, K. O'Connor and R. Li *et al.*, 2023. ChatGPT and large language models in academia: Opportunities and challenges. *Biodata Min.*, Vol. 16. 10.1186/s13040-023-00339-9.
29. Lund, B.D., T. Wang, N.R. Mannuru, B. Nie, S. Shimray and Z. Wang, 2023. ChatGPT and a new academic reality: Artificial Intelligence-written research papers and the ethics of the large language models in scholarly publishing. *Assoc. Inf. Sci. Technol.*, 74: 570-581.
30. Miller, T., 2019. Explanation in artificial intelligence: Insights from the social sciences. *Artif. Intell.*, 267: 1-38.
31. Funnell, E.L., B. Spadaro, J. Benacek, N.A. Martin-Key and T. Metcalfe *et al.*, 2022. Learnings from user feedback of a novel digital mental health assessment. *Front. Psychiatry*, Vol. 13. 10.3389/fpsyt.2022.1018095.
32. Resnik, D.B. and M. Hosseini, 2025. The ethics of using artificial intelligence in scientific research: New guidance needed for a new tool. *AI Ethics*, 5: 1499-1521.
33. O'Connor, S. and ChatGPT, 2023. Open artificial intelligence platforms in nursing education: Tools for academic progress or abuse? *Nurse Educ. Pract.*, Vol. 66. 10.1016/j.nepr.2022.103537.
34. Page, M.J., J.E. McKenzie, P.M. Bossuyt, I. Boutron and T.C. Hoffmann *et al.*, 2021. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, Vol. 372. 10.1136/bmj.n71.
35. Rahman, M.M. and Y. Watanobe, 2023. ChatGPT for education and research: Opportunities, threats, and strategies. *Appl. Sci.*, Vol. 13. 10.3390/app13095783.
36. Mizanur Rahman, H.J.R. Terano, Nafizur Rahman, A. Salamzadeh and Saidur Rahaman, 2023. ChatGPT and academic research: A review and recommendations based on practical examples. *J. Educ. Manage. Dev. Stud.*, 3: 1-12.
37. Doskaliuk, B., O. Zimba, M. Yessirkepov, I. Klishch and R. Yatsyshyn, 2025. Artificial intelligence in peer review: Enhancing efficiency while preserving integrity. *J. Korean Med. Sci.*, Vol. 40. 10.3346/jkms.2025.40.e92.
38. Farber, S., 2025. Comparing human and AI expertise in the academic peer review process: Towards a hybrid approach. *Higher Educ. Res. Dev.*, 44: 871-885.
39. Yang, Z., X. Zhou, Y. Jiang, X. Zhang, Q. Gao, Y. Lu and A. Yang, 2026. Human-AI complementarity in peer review: Empirical analysis of PeerJ data and design of an efficient collaborative review framework. *Publications*, Vol. 14. 10.3390/publications14010001.
40. Stokel-Walker, C., 2023. ChatGPT listed as author on research papers: Many scientists disapprove. *Nature*, 613: 620-621.
41. Supheert, R.T., J. van Dillen and C. Fluit, 2023. Teaching medical students to teach: Supplementing a narrative review with gray literature. *Acad. Med.*, 98: 976-976.
42. Thorp, H.H., 2023. ChatGPT is fun, but not an author. *Science*, 379: 313-313.