https://doi.org/10.70844/ijas.2026.3.57

Innovative Journal of Applied Science

Review Article ISSN: 3049-3072

IJAS-26-151

Responsible AI in Academic Research: A Scoping Review and Evidence-Gap Map of Ethical Principles, Challenges and Governance

Fredrick M. Ogore*

Department of Computer Science, School of Science, Engineering and Health, Daystar University, Nairobi, Kenya

Corresponding Author: Fredrick M. Ogore, Department of Computer Science, School of Science, Engineering and Health, Daystar University, Nairobi, Kenya, E-mail: fogore@daystar.ac.ke

Received date: 31 October, 2026, **Accepted date:** 16 November, 2026, **Published date:** 24 November, 2026

Citation: Ogore FM (2026) Responsible AI in Academic Research: A Scoping Review and Evidence-Gap Map of Ethical Principles, Challenges and Governance. Innov J Appl Sci 3(6): 57.

Abstract

Introduction: Artificial Intelligence (AI) is now used at every stage of the academic research life cycle, yet the rules governing its use remain fragmented and the evidence base is dispersed across disciplines and document types.

Objective: To map the nature, extent and distribution of evidence on the ethics of using AI to conduct, report, review and govern academic research and to identify where evidence and governance are lacking.

Methods: A scoping review was conducted following JBI guidance and reported according to the PRISMA extension for scoping reviews (PRISMA-ScR). Thirty-four structured search strings and citation chasing were used to identify documents published between 2015 and 2025. Each principal finding was linked to the type and amount of supporting evidence using a transparent confidence rule and included items were cross-tabulated by theme and research-lifecycle stage in an evidence-gap map.

Results: Of 359 records identified, 124 met the eligibility criteria: 40 empirical studies, 43 conceptual analyses or commentaries, 20 policy and guidance documents, 17 reviews and 4 methodological frameworks; 83% were published from 2023 onwards. High-confidence findings were that guidance converges on no AI authorship, disclosure of substantive use and human accountability; that generative-AI use in research writing is widespread and largely undisclosed; that large language models fabricate references; and that AI-text detectors are unreliable and biased. Evidence and guidance were concentrated at the writing (76 items) and peer-review (39 items) stages, whereas the study-design and data stages were each supported by only four empirical studies.

Conclusions: Ethical attention and governance need to move upstream to the design and data stages of research. An evidence-weighted lifecycle framework is proposed that assigns governance gaps to researchers, institutions, publishers, funders and regulators, together with priorities for upstream empirical research, ethics-review reform, AI literacy in research training and international coordination.

Keywords: Artificial intelligence, Research ethics, Research integrity, Generative AI, Large language models, Scoping review, Evidence-gap map, Research governance, Higher education

Introduction

Artificial Intelligence (AI) has moved from a specialized research object to a general-purpose research instrument. Machine learning models now predict protein structures with near-experimental accuracy, accelerate hypothesis generation and experimental design across the natural sciences and, since the public release of Large Language Models (LLMs) such as ChatGPT in late 2022, assist researchers with literature searching, coding, drafting, translation and editing [1,2]. A survey of more than 1,600 researchers conducted by Nature found that a substantial share already used AI tools in their work and expected those tools to become central to research practice within a decade, while simultaneously voicing concern about misinformation, fraud and bias [3]. Large-scale linguistic analyses indicate that LLM-modified text rose rapidly in scientific abstracts after 2022, most steeply in computer science [4].

This diffusion has outpaced the evolution of common ethical standards. Traditional research-ethics frameworks, such as the principles of beneficence, justice and respect for persons in the Belmont Report, were developed for investigator-participant relationships in which the researcher had control over the instruments

of inquiry and could provide an explanation [5]. AI violates these assumptions: models can be black boxes to their designers, trained on data collected without the knowledge of the people they describes and capable of producing fluent but false content. At the same time, the multiplication of AI ethics guidelines, which had surpassed eighty by 2019, has resulted in a gap between abstract principles and operational practice, as noted by several authors [6-8].

Existing reviews have mapped general AI ethics principles, the use of generative AI in education and particular question of AI authorship [6,9-11]. Few studies collate the ethical consequences of using AI for academic research as a practice including responses of researchers, institutions, publishers and regulators, areas of agreement and disagreement and remaining unresolved challenges in one framework. Resnik and Hosseini contend that current guidance on research ethics needs to be revised considerably to include AI as a research tool [12]. But systematic evidence on current practices is dispersed across disciplines and document types.

The question is especially pressing for higher education. Universities are at once the main producers of academic research, the institutions responsible for research-integrity oversight through ethics committees and misconduct procedures and the places where the next generation of researchers is trained. Analyses of institutional policies show that universities have issued guidance on generative AI rapidly, but with considerable variation in how research uses are addressed [13]. A clear map of the ethical evidence is therefore needed not only by publishers and regulators, but also by universities designing research-integrity policies, ethics-review procedures and research-methods curricula.

This scoping review addresses that gap. A scoping design was chosen because the field is new, fast-moving and methodologically heterogeneous and because the aim is to map the nature, extent and distribution of the evidence rather than to estimate effects or answer a narrowly defined question [14,15]. Drawing on primary sources (empirical studies, audits and surveys) and secondary sources (conceptual analyses, reviews and policy and guidance documents), it asks five research questions (RQs):

RQ1. What fundamental principles of ethical research are invoked for the use of AI in academic research and how are they interpreted in this context?

RQ2. What are the current practices of researchers, institutions, publishers and regulators and where do they converge or diverge?

RQ3. What are the principal ethical challenges associated with AI-enabled research?

RQ4. What strategies are emerging to address these challenges and what future research is needed?

RQ5. How are evidence and guidance distributed across the stages of the research lifecycle and where are the principal gaps?

The review makes three contributions. First, it maps the research-specific evidence on the ethics of AI in academic research, distinguishes this reviewed corpus from the broader AI-ethics literature used only as conceptual background and links each principal finding to the type and amount of evidence that supports it. Second, it provides a structured comparison of where publisher, editorial, institutional and regulatory guidance converges and diverges. Third, it presents an evidence-gap map across six research-lifecycle stages and, on that basis, an evidence-weighted framework that differs from existing lifecycle approaches by showing where evidence and governance are thin and which actors should close each gap. The paper then sets out the conceptual background, the methods, the results and the discussion, and closes with a summary and conclusions.

Conceptual background: Principles of ethical research in the age of AI

This section draws on general research-ethics and AI-ethics scholarship to set out the principles that frame the analysis. These sources provide conceptual background and are not part of the reviewed corpus described in in the Methodology; in the reference list, studies included in the scoping review are marked with an asterisk.

Historically, research ethics has been founded on principlism, which involves the application of a limited number of mid-level principles to organize moral reasoning [16]. A comparable framework has been implemented in the field of AI ethics. Floridi and Cowls [19]; proposed a unified framework that would enhance the explicability of the four bioethical principles, while Jobin et al. [6] identified transparency, justice and fairness, non-maleficence, responsibility and privacy as the most frequently cited principles across 84 guidelines [6]. The OECD Recommendation on AI and the UNESCO Recommendation on the Ethics of Artificial Intelligence, which was adopted by its 193 member states, are intergovernmental instruments that codify comparable commitments [17,18]. Five principles were identified as the most pertinent to academic research (RQ1) for the purposes of this review.

Transparency and explicability

Research transparency has historically required methods to be described in a way that is sufficiently detailed to permit examination and replication. AI transparency has three layers: disclosure (if and how AI was used), documentation (the data, model, parameters and prompts) and explicability (if the reasoning behind model outputs can be understood). The third layer is technical tools for explainable AI research [20]. For high-stakes decisions, Rudin argues that researchers should give priority to models that are naturally interpretable rather than post-hoc explanations of black-box systems [21]. Additionally, reproducibility depends on openness: According to Haibe-Kains et al., high-profile AI studies may become effectively unverifiable due to the lack of shared code and model details [22].

https://doi.org/10.70844/ijas.2026.3.57

Innovative Journal of Applied Science

Justice and fairness

The principle of justice relates to the fair allocation of the benefits and burdens of research. In AI-enabled research, this includes algorithmic fairness, which is the degree to which models trained on biased data systematically disadvantage groups based on race, gender, language, geography, or socioeconomic status [23]. The structural dimension of justice includes data extraction, training corpora knowledge system representation and access to computational resources [24,25].

Respect for persons, autonomy and privacy

Respect for individuals underpins informed consent and the protection of individual rights. These mechanisms are often contested in big-data and AI research, as data is often repurposed, scraped from public platforms or combined in ways that participants could not have foreseen [26]. Rocher et al. estimated that 99.98% of Americans can be re-identified correctly in any dataset using 15 demographic attributes, suggesting that de-identification offers weaker protection than is commonly assumed [27].

Accountability and responsibility

Responsibility implies that identifiable individuals should be accountable for the conduct and outcomes of research. Publishers and ethics bodies have clarified that artificial intelligence systems are not able to meet authorship criteria because they lack the ability to assume moral or legal obligations [28,29]. Accountability requires that decisions based on AI are contestable and responsibility is not spread over developers, vendors and users [30].

Research integrity and beneficence

When generative models produce plausible but fictitious content, research integrity, which encompasses honesty, rigor and the avoidance of fabrication, falsification and plagiarism, is directly engaged. AI-enabled research should also be beneficent and non-maleficent, that is, it should offer a net benefit and avoid foreseeable harm, including environmental costs associated with training large models [31,32].

Table 1 summarizes how these principles are operationalized into specific duties across the research life-cycle. It is also the analytical lens used during data charting (see Data charting and synthesis).

Principle	Core question for AI-enabled research	Illustrative operational obligations
Transparency / explicability	Can others understand what AI did and verify it?	Disclose AI tools, versions and purposes; share code, data and prompts where lawful; prefer interpretable models in high-stakes settings
Justice / fairness	Who benefits, who is burdened and is the model biased?	Audit training data and outputs for subgroup performance; report demographic composition; ensure equitable access and benefit sharing
Respect for persons / privacy	Were data subjects' rights and expectations respected?	Obtain appropriate consent or ethics waiver; minimise and secure data; assess re-identification risk; avoid uploading confidential data to third-party tools
Accountability	Who answers for AI-assisted outputs and decisions?	Only humans as authors; named responsibility for verification; documented human oversight; mechanisms for contestation and correction
Integrity / beneficence	Are outputs genuine and is the research net beneficial?	Verify every AI-generated claim and citation; prohibit AI-fabricated data and images; weigh environmental and social costs

Table 1: Operationalization of core ethical principles for AI-enabled academic research. **Source:** Author's synthesis based on Beauchamp and Childress [16]; Floridi and Cowls [19]; Jobin et al. [6] and UNESCO [18].

Methodology

Review design

This review followed the JBI methodology for scoping reviews building on the framework of Arksey and O'Malley as refined by Levac et al. and is reported according to the PRISMA extension for scoping reviews (PRISMA-ScR) [15,33-35]. A scoping design was preferred to a systematic review because the evidence base is heterogeneous, combining empirical, conceptual, normative and policy documents and because the purpose was to map concepts, evidence types and gaps rather than to synthesise effect estimates [14].

Information sources and search strategy

Searches were run on 14 October 2026 and eligibility was restricted to documents published between 1 January 2015 and 31 December 2025. Exports from subscription bibliographic databases (Scopus, Web of Science) were not available, so the search was designed as a structured, multi-string search of a general scholarly web search engine that indexes publisher platforms (including ScienceDirect, SpringerLink, Wiley Online Library, SAGE and Taylor & Francis), PubMed/PMC, the ACM Digital Library and the arXiv and medRxiv preprint servers. Thirty-four search strings were constructed, each combining a technology concept (e.g., artificial intelligence, large language models, ChatGPT, generative AI) with an ethics or integrity concept (e.g., ethics, integrity, bias, privacy, transparency, disclosure, authorship) and a research-practice context (e.g., scientific research, scholarly publishing, peer review, ethics committees); the set was designed so that every theme in the protocol was targeted by at least one string. Retrieval was limited to the first page of ranked results for each string (9–10 records per string; 314 in total) and every record returned was logged with its URL (Supplementary File S1, sheets Search_Log and Screening_Log).

https://doi.org/10.70844/ijas.2026.3.57

Innovative Journal of Applied Science

Because search-engine ranking is not fully deterministic, an identical re-run may return a partly different set of records; the strings, run date and complete record log are therefore reported so that the search can be audited and re-executed in bibliographic databases. To offset the retrieval limit, the search was supplemented by backward citation chasing from included studies and by known-item checks of guidance issued by COPE, ICMJE, WAME, major journals and publishers, the European Union and UNESCO. The generic Boolean structure underlying the strings was:

("artificial intelligence" OR "machine learning" OR "deep learning" OR "large language model*" OR "generative AI" OR ChatGPT) AND (ethic* OR integrity OR "responsible research" OR bias OR fairness OR transparen* OR privacy OR accountab*) AND ("academic research" OR "scientific research" OR "scholarly publishing" OR "research practice*" OR "peer review" OR authorship OR "research ethics")

Eligibility criteria

Eligibility criteria were defined using an adapted Population–Concept–Context framework suited to scoping and qualitative reviews (Table 2).

Dimension	Inclusion	Exclusion	Code
Publication date	1 January 2015 – 31 December 2025	Published before 2015 or after 2025	E1
Population / actors	Researchers, research institutions, ethics committees/IRBs, funders, publishers, editors, reviewers, regulators	Consumers, industry product teams, clinicians in routine care, students in classroom assessment	E2
Concept	Ethical, integrity, or governance aspects of using AI to conduct, report, review, or govern research	Purely technical or how-to work without ethical or integrity analysis	E3
Context	Academic and scientific research in any discipline, including scholarly publishing and peer review	General AI ethics without a research-practice component; clinical deployment; industry practice; teaching and learning	E2
Document type	Peer-reviewed empirical, review and conceptual articles; peer-reviewed commentaries; preprints from recognised servers reporting original data or analysis; official guidance from publishers, editorial bodies, governments and intergovernmental organisations	News, blogs, vendor pages, social media, library how-to guides, slides, AI-generated summaries	E4
Verifiability	Venue and authorship verifiable	Unrefereed uploads without venue; case notes; unindexed outlets	E5
Language	English	Other languages	E2
Insufficient information	–	Title, date, or venue could not be verified (reported as awaiting classification)	AWC

Table 2: Eligibility criteria.

Reviewed corpus and contextual literature. Only documents meeting all criteria form the reviewed corpus on which the Results are based. Broader studies of algorithmic fairness, of AI deployment in health care or commerce and general AI-ethics guidelines were excluded from the corpus (code E2) but are cited in the Introduction, Conceptual background and the discussion. Where such sources appear in the Results, they are explicitly labelled as contextual and are not counted as evidence for any finding.

Selection of sources of evidence

All records were logged in a screening sheet (Supplementary File S1) and duplicates, including alternative copies, versions and aggregator listings of the same work, were removed. Titles and abstracts were screened against the eligibility criteria and potentially eligible reports were assessed in full where accessible. Each exclusion was assigned a reason code (E1–E5; Table 2) and records whose bibliographic details could not be verified were reported as awaiting classification.

Data charting and synthesis

Data were charted using a standardised form capturing bibliographic details, document type, discipline, research-lifecycle stage addressed (design, data, analysis, writing, peer review, governance), ethical principles invoked (Table 1), themes and key findings (Supplementary File S1, sheet Included_Studies). For 72 of the 124 included items (58%), charting was based on the full text; for the remaining 52 (42%), it was based on the title and abstract and these entries are flagged in the workbook. Results were synthesised through a descriptive numerical summary and qualitative content analysis organised by research question, following JBI guidance [15]. For RQ2, guidance documents were additionally compared on a common set of provisions.

Evidence weighting and evidence-gap mapping

Consistent with scoping-review methodology, no formal risk-of-bias appraisal of individual studies was performed and earlier plans to apply the MMAT and AACODS tools were not pursued because

many included items are conceptual or normative [33]. Instead, to make explicit how different kinds of evidence contribute to the conclusions, every included item was classified by evidence type (empirical study, review, policy or guidance document, conceptual analysis or commentary, methodological framework) and each principal finding was linked to the specific included items that support it (Table 3). A transparent rule was used to express confidence in each finding: high, when supported by at least three empirical studies with concordant results or at least five concordant authoritative guidance documents; moderate, when supported by one or two empirical studies or by two to four concordant guidance documents; and emerging, when supported mainly by conceptual analyses or commentaries. The rule describes the breadth and type of support, not the internal validity of individual studies. Finally, included items were cross-tabulated by theme and lifecycle stage to produce an evidence-gap map, with empirical studies counted separately [37].

Principal finding	Supporting included studies (IDs)	n (empirical / guidance / other)	Charted from full text	Confidence
Guidance converges on three norms: no AI authorship, disclosure of substantive use and full human accountability	S088, S089, S091, S092, S093, S094, S119, S065, S081, S003, S072, S073, S074, S013	14 (5 / 9 / 0)	11 of 14	High
Disclosure thresholds and formats diverge across policies and in the normative literature	S048, S049, S007, S018, S072, S073	6 (3 / 0 / 3)	4 of 6	High
Generative-AI use in research writing is widespread and largely undisclosed	S087, S101, S079, S103, S019, S083, S018	7 (7 / 0 / 0)	4 of 7	High
LLMs fabricate or distort references and factual claims	S097, S022, S023, S098, S024, S025	6 (3 / 0 / 3)	6 of 6	High
AI-text detectors are unreliable and biased against non-native English writers	S026, S027, S100, S099	4 (4 / 0 / 0)	4 of 4	High
AI-generated research images are difficult for experts to detect	S056, S057, S055	3 (1 / 0 / 2)	1 of 3	Moderate
LLM use in peer review is already common and threatens confidentiality and fairness	S042, S041, S012, S104, S063, S011, S010	7 (4 / 0 / 3)	3 of 7	High
Ethics committees lack the mandate and expertise to review AI-specific research risks	S020, S021, S105, S106, S107, S029	6 (0 / 1 / 5)	5 of 6	Emerging
AI-enabled research methods can embed bias (e.g., LLMs as simulated participants)	S112, S122, S121, S123, S028, S100	6 (2 / 0 / 4)	5 of 6	Moderate
Leakage and incomplete reporting undermine reproducibility of ML-based research	S031, S109, S110, S032, S030	5 (2 / 0 / 3)	3 of 5	Moderate
Documentation and reporting tools exist but uptake is partial	S046, S044, S045, S111, S114, S115, S116	7 (1 / 3 / 3)	7 of 7	Moderate
Research-community self-governance (broader-impact statements) is uneven	S068, S069, S070	3 (3 / 0 / 0)	3 of 3	High
Equity, Global South participation and public involvement are under-studied	S043, S053, S054, S058, S076	5 (2 / 0 / 3)	0 of 5	Moderate
Epistemic and environmental risks of AI in science are raised mainly conceptually	S113, S124, S102, S085, S067, S066	6 (2 / 0 / 4)	5 of 6	Moderate

Table 3: Principal findings, supporting evidence from the reviewed corpus and confidence ratings. **Note:** Study IDs refer to Supplementary File S1 (sheet Included_Studies). Confidence: high = at least three concordant empirical studies or at least five concordant authoritative guidance documents; moderate = one or two empirical studies or two to four guidance documents; emerging = mainly conceptual. Ratings describe the breadth and type of support, not study quality. Scope caveats: Evidence on peer review comes mainly from machine-learning conferences and evidence on detectors mainly from tests on educational and biomedical texts.

Use of AI tools

Quill Bot, Grammarly and Paper pal were used for language grammar and editing. All AI outputs were reviewed by the author who takes full responsibility for the content. No AI tool is credited as an author. Publish or perish retrieving and analysing academic citations.

Results

Characteristics of the included evidence

The searches identified 314 records and a further 45 were identified through other methods (Figure 1). After removal of 63 duplicates, 296 records were screened by title and abstract, of which

https://doi.org/10.70844/ijas.2026.3.57

Innovative Journal of Applied Science

139 were excluded: 37 fell outside the 2015–2025 window (predominantly items published in 2026), 55 were non-scholarly sources and 47 addressed AI ethics outside academic research practice. Of the 157 reports assessed in full, 5 lacked an ethical or integrity focus, 9 were neither peer-reviewed nor authoritative guidance and 19 could not be verified and are reported as awaiting classification. In total, 124 studies and documents were included in the synthesis.

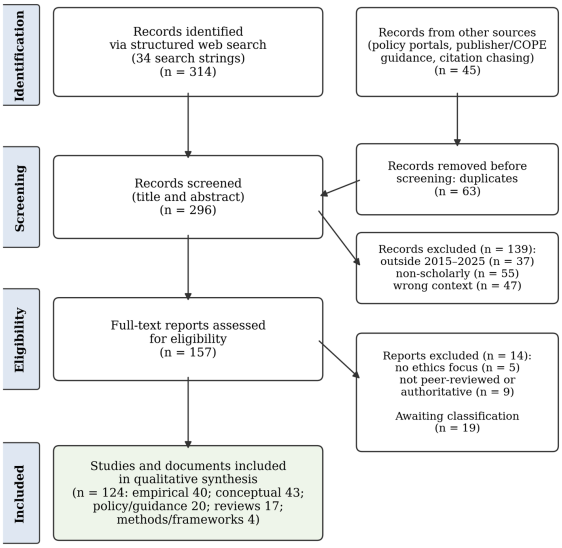

Figure 1: Flow diagram of the identification, screening and inclusion of sources of evidence (adapted from PRISMA 2020) [36]. Searches were run on 14 October 2026; publication window 2015–2025.

The included evidence is recent and mostly normative (Figure 2). Only 21 items (17%) were published between 2015 and 2022, whereas the remaining 103 (83%) appeared in 2023–2025, after the public release of ChatGPT. The most common are conceptual analyses and commentaries (n = 43) and empirical studies (n = 40), followed by policy and guidance documents (n = 20), reviews (n = 17) and methodological frameworks (n = 4). Empirical work has concentrated on four designs: corpus and linguistic analyses to estimate the prevalence of AI-generated text [4,38,39]; experimental tests of generative-AI outputs and detectors [40-42]; analyses of journal and institutional policies [43,44]; and researcher surveys [3]. Publication policy (n = 28 studies), research integrity (n = 26) and authorship and disclosure (n = 19) were the most frequently addressed themes, whereas global equity (n = 4), environmental sustainability (n = 2) and public participation (n = 1) were the least studied. Medicine and the biomedical sciences, computer science and research-ethics scholarship contributed most of the evidence. All 124 items address AI in research practice by design; the broader literature cited in Section 2 as conceptual background is not included in these counts.

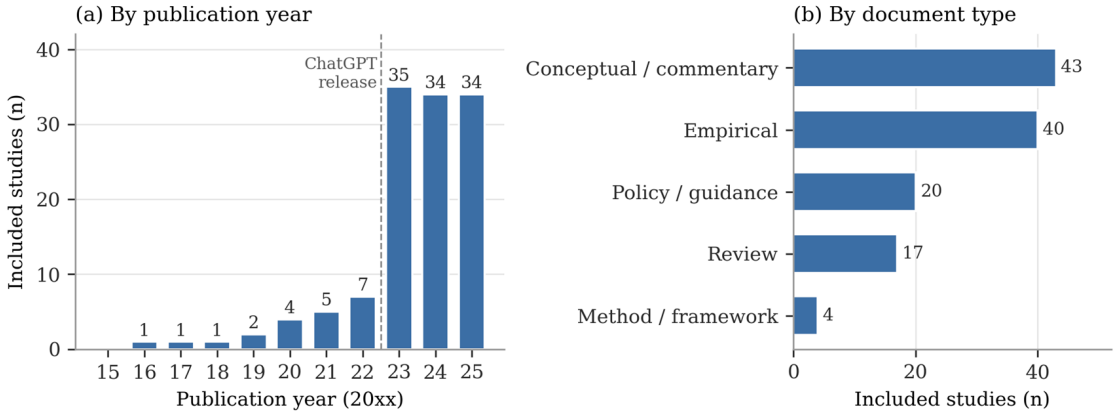

Figure 2: Included studies (n = 124) by (a) publication year and (b) document type. The dashed line marks the public release of ChatGPT (November 2022).

https://doi.org/10.70844/ijas.2026.3.57

Innovative Journal of Applied Science

Evidence-gap map (RQ5)

Figure 3 cross-tabulates the included items by theme and lifecycle stage. Evidence is heavily concentrated at the writing stage (76 items, including 26 empirical studies) and, to a lesser extent, at peer review (39 items; 13 empirical) and governance (38 items; 12 empirical). By contrast, the design and data stages are addressed by 23 and 29 items respectively, but by only four empirical studies each and these are general surveys of AI use rather than studies of design- or data-stage practice. Guidance follows the same pattern: of the 20 included policy and guidance documents, 13 address writing and 7 peer review, whereas only 3 address study design and 4 data collection, mostly through general provisions. Three cells stand out. Integrity of outputs at the writing stage is the most densely studied area (34 items; 8 empirical). Data governance and ethics review at the data stage is addressed by eight items, none of them empirical. Bias and fairness appears at every stage but is addressed by only two empirical studies in total and equity and participation by five items, two of them empirical.

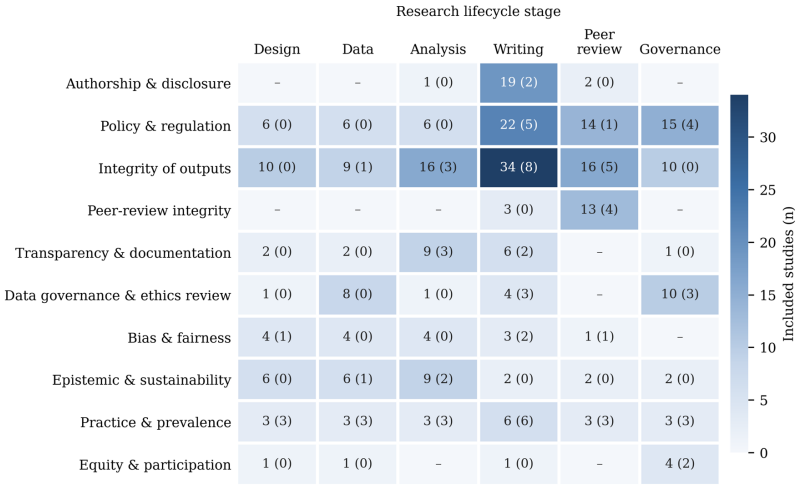

Figure 3: Evidence-gap map of the included studies (n = 124) by theme and research-lifecycle stage. Cell values give the number of included items, with the number of empirical studies in parentheses; an item can contribute to several cells. Items addressing all stages (e.g., general surveys of AI use) are counted in every stage.

Current practices: Convergence and divergence (RQ2)

Areas of convergence

Guidance issued by editorial bodies and publishers following the emergence of ChatGPT converged rapidly on three norms. First, AI tools cannot be authors. After ChatGPT was listed as a co-author on several papers and preprints, Nature and Science declared that LLMs do not satisfy authorship criteria, a position subsequently adopted by COPE, WAME, ICMJE, the JAMA Network, a group of bioethics journal editors and Elsevier [28,29]. Analyses of author instructions confirm the pattern: By 2023, 87% of the top 100 journals but only 24% of the top 100 publishers provided guidance on generative AI and nearly all that did prohibited AI authorship [43] (see also [44]). This is because authorship entails responsibility and a non-agent cannot be responsible. Second, any substantive use of AI should be disclosed, generally either in the methods or acknowledgments section or as a dedicated statement. Third, the accuracy, originality and integrity of all content, including any produced with AI assistance, are the sole responsibility of human researchers [45-52].

Outside research-specific guidance, general AI-governance instruments show a similar convergence. The commitments to transparency, fairness, human oversight, privacy and accountability in the UNESCO Recommendation, which forms part of the corpus, are shared by the OECD principles and the NIST AI Risk Management Framework, which are cited here as context and echo the principle-level consensus documented by Jobin et al. [6,17,18,53].

Areas of divergence

Beneath this consensus, practice varies in at least five respects (Table 4). (i) Disclosure thresholds vary: some policies exclude basic language editing and grammar correction, while others require disclosure of any generative AI use. The normative literature is split. Resnik and Hosseini argue for mandatory disclosure when AI plays a substantive role in shaping content, data, or analysis, but optional disclosure for routine language editing. Disclosure of writing assistance is at the discretion of the author, according to Hosseini et al. Science prohibited AI-generated text outright at first, before adopting a disclosed use policy. That shows how quickly the mood has changed. (ii) Peer review is viewed in an inconsistent way. Some organizations prohibit reviewers from uploading manuscripts to generative-AI tools due to confidentiality concerns and researchers have suggested carefully limited uses to combat reviewer burnout [29,52]. At the same time, this usage is already prevalent, with estimates that LLMs significantly altered 6.5–16.9% of peer-review text at major AI conferences in 2023–2024, AI-assisted reviews were

associated with higher scores and acceptance rates and existing policies have been criticized as inadequate to safeguard review integrity [54-57]. (iii) Publishers generally do not permit use of AI-generated images and data unless they are part of the research design, but institutional rules vary. (iv) Some institutions employ detection tools despite evidence of their unreliability and bias against non-native English writers, misclassifying more than half of TOEFL essays as AI-generated [41,58,59]. (v) Regulatory scope varies. The EU Artificial Intelligence Act significantly excludes artificial intelligence systems developed and used exclusively for scientific research and development from its scope, deferring research governance to institutional and professional self-regulation [60-62].

Source	AI as author	Disclosure required	AI in peer review	Distinctive provisions
COPE [28]	Not permitted	Yes	Not specified in position statement	Authors fully responsible for content, including AI-produced parts
ICMJE [29]	Not permitted	Yes (cover letter and manuscript)	Reviewers must not upload manuscripts where confidentiality cannot be assured	Authors should be able to assert no plagiarism in AI-produced text and images
WAME [47]	Not permitted	Yes	Disclosure required if used	Editors need appropriate tools to detect AI-generated content
Nature [51]	Not permitted	Yes (methods or acknowledgements)	Restricted	Generally, does not permit AI-generated images
Science [46]	Not permitted	Initially banned text; later disclosed use permitted	Restricted	Policy revised within the same year
Elsevier [52]	Not permitted	Yes (dedicated declaration)	Reviewers and editors must not upload manuscripts	AI-generated images not permitted unless part of research design
JAMA Network [48]	Not permitted	Yes (acknowledgements or methods)	Must not enter manuscripts into tools that breach confidentiality	Authors responsible for accuracy of AI-generated content
European Commission [92]	n/a	Yes (substantial use)	Avoid substantial use in peer review and evaluation	Living guidelines for researchers, research organisations and funders
EU AI Act [60]	n/a	Transparency duties for certain systems	n/a	Exclusion for AI used solely for scientific R&D
UNESCO [18]; OECD [17]; NIST [53]	n/a	Transparency as general principle	n/a	Principle-based and voluntary; not research-specific

Table 4: Convergence and divergence in selected guidance on AI use in research and scholarly publishing. **Note:** Policies evolve rapidly; entries reflect the cited versions and should be checked against current publisher and regulator guidance.

Researcher and institutional practice

From an empirical point of view, practice has led policy. Surveys indicate widespread acceptance of generative AI for writing and coding and literature work, with uncertainty about disclosure and concern about accuracy [3,63]. Linguistic analyses suggest that use is massively under-disclosed: about 1% of papers published in 2023 contained marker vocabulary of LLM assistance [38]; LLM-modified content made up as much as 17.5% of computer-science abstracts by early 2024 [4]; and at least 13.5% of 2024 PubMed abstracts were processed with LLMs well beyond rates of formal disclosure [39]. Often, institutional policies are delegated to the faculty or departmental level, resulting in inconsistency across and within universities. Research ethics committees and IRBs do not typically have the mandate or expertise to review AI-specific risks including dataset provenance, model bias and downstream societal harms, especially for research that uses existing or publicly available data and is thus outside of traditional human-subjects review [26,64-66]; in response, the U.S. Secretary's Advisory Committee on Human Research Protections released specific considerations for IRB review of AI research [67]. The NeurIPS broader-impact requirement provided a natural experiment in self-governance in computer science, but analyses found statements of uneven depth that rarely engaged with concrete harms [68-70].

Principal ethical challenges (RQ3)

Content analysis identified three dominant challenge clusters and two cross-cutting concerns. Table 3 links each principal finding to the included studies that support it and to the resulting confidence rating; the evidence-gap map (Figure 3) shows where these studies are located in the research lifecycle.

https://doi.org/10.70844/ijas.2026.3.57

Innovative Journal of Applied Science

Principal finding	Supporting included studies (IDs)	n (empirical / guidance / other)	Charted from full text	Confidence
Guidance converges on three norms: no AI authorship, disclosure of substantive use and full human accountability	S088, S089, S091, S092, S093, S094, S119, S065, S081, S003, S072, S073, S074, S013	14 (5 / 9 / 0)	11 of 14	High
Disclosure thresholds and formats diverge across policies and in the normative literature	S048, S049, S007, S018, S072, S073	6 (3 / 0 / 3)	4 of 6	High
Generative-AI use in research writing is widespread and largely undisclosed	S087, S101, S079, S103, S019, S083, S018	7 (7 / 0 / 0)	4 of 7	High
LLMs fabricate or distort references and factual claims	S097, S022, S023, S098, S024, S025	6 (3 / 0 / 3)	6 of 6	High
AI-text detectors are unreliable and biased against non-native English writers	S026, S027, S100, S099	4 (4 / 0 / 0)	4 of 4	High
AI-generated research images are difficult for experts to detect	S056, S057, S055	3 (1 / 0 / 2)	1 of 3	Moderate
LLM use in peer review is already common and threatens confidentiality and fairness	S042, S041, S012, S104, S063, S011, S010	7 (4 / 0 / 3)	3 of 7	High
Ethics committees lack the mandate and expertise to review AI-specific research risks	S020, S021, S105, S106, S107, S029	6 (0 / 1 / 5)	5 of 6	Emerging
AI-enabled research methods can embed bias (e.g., LLMs as simulated participants)	S112, S122, S121, S123, S028, S100	6 (2 / 0 / 4)	5 of 6	Moderate
Leakage and incomplete reporting undermine reproducibility of ML-based research	S031, S109, S110, S032, S030	5 (2 / 0 / 3)	3 of 5	Moderate
Documentation and reporting tools exist but uptake is partial	S046, S044, S045, S111, S114, S115, S116	7 (1 / 3 / 3)	7 of 7	Moderate
Research-community self-governance (broader-impact statements) is uneven	S068, S069, S070	3 (3 / 0 / 0)	3 of 3	High
Equity, Global South participation and public involvement are under-studied	S043, S053, S054, S058, S076	5 (2 / 0 / 3)	0 of 5	Moderate
Epistemic and environmental risks of AI in science are raised mainly conceptually	S113, S124, S102, S085, S067, S066	6 (2 / 0 / 4)	5 of 6	Moderate

Table 3: Principal findings, supporting evidence from the reviewed corpus and confidence ratings. **Note:** Study IDs refer to Supplementary File S1 (sheet Included_Studies). Confidence: high = at least three concordant empirical studies or at least five concordant authoritative guidance documents; moderate = one or two empirical studies or two to four guidance documents; emerging = mainly conceptual. Ratings describe the breadth and type of support, not study quality. Scope caveats: evidence on peer review comes mainly from machine-learning conferences and evidence on detectors mainly from tests on educational and biomedical texts.

Algorithmic bias and representational harm

Within the reviewed corpus, evidence on bias in AI-enabled research is limited and mostly conceptual: across all lifecycle stages, only two included empirical studies address bias directly (Figure 3). Both concern the infrastructure of research rather than its findings. AI-text detectors used to police research writing misclassify non-native English writing as machine-generated and an analysis of 100 highly cited machine-learning papers found that values such as performance and generalisation dominate, whereas societal needs and potential harms are rarely discussed [59,71]. A newer concern, raised in perspective articles, arises when LLMs are used as substitutes for human participants or as research assistants in the social and behavioural sciences: model outputs reflect the skewed populations and viewpoints represented in training data, so findings may not generalise to under-represented groups [72-74]. As contextual background from outside research settings, audits of deployed systems show how such biases arise, for example through unrepresentative training data or cost-based proxies for need [75,76] (see also [23]); when similar tools are used as research instruments, bias becomes a validity problem as well as an ethical one. Selbst et al., warn that purely technical fairness fixes may fail if they abstract away from social context [77].

Data governance, privacy and consent

The reviewed literature repeatedly notes that the provenance, consent status and licensing of the datasets underpinning AI research are often unclear and that the public availability of web-scraped data is not equivalent to consent for research use [26,65,78]. Contextual evidence on re-identification further questions the adequacy of anonymisation as a safeguard [27]. A challenge specific to generative AI is the transmission of confidential content, such as unpublished manuscripts, participant data and grant proposals, to third-party services whose data-retention and training practices may be opaque; this underlies the confidentiality provisions in peer-review policies [29,52]. Notably, none of the eight included items addressing data governance at the data stage is empirical (Figure 3): the risks are well argued, but their prevalence and consequences in research practice remain unmeasured. Robust data governance aligned with the FAIR

principles must therefore be reconciled with privacy and data-protection obligations [79].

Transparency, reproducibility and research integrity

Insufficient reporting and ambiguity threaten reproducibility. Gundersen and Kjensmo found that only a small portion of AI papers include enough information for replication. Kapoor and Narayanan discovered data leakage in 294 papers across 17 scientific fields using machine learning which inflated the reported performance. Generative AI brings new integrity risks. LLMs "hallucinate" plausible but false statements and references. Walters and Wilder found that 55% of citations generated by GPT-3.5 and 18% by GPT-4 were fabricated [40] (see also [80]). Comparable results were reported in medicine, where nearly half of the references in ChatGPT-generated medical content were fabricated and most of the remainder inaccurate and hallucination rates for references proposed for systematic reviews reached 39.6% for GPT-3.5, 28.6% for GPT-4 and 91.4% for Bard [81,82]. Gao et al. showed that blinded human reviewers identified only 68% of ChatGPT-generated scientific abstracts, while misclassifying 14% of genuine abstracts as generated, which highlights the potential for undetectable paper-mill content [83]. Image integrity is similarly threatened: generative models can produce realistic, fabricated research images and expert pathologists could not reliably distinguish AI-generated from genuine histological images [42,84]. Messeri and Crockett further warn of "illusions of understanding," in which reliance on AI leads scientists to believe they understand more than they do and narrows the range of questions pursued, a concern echoed in expert assessments of how LLMs should and should not change scientific practice [85-89].

Cross-cutting concerns: Accountability and sustainability

Two further concerns recur. First, diffuse accountability: when AI assists with data analysis or writing, it may be unclear whether researchers, tool developers, or institutions are responsible for errors, a problem also analysed in the context of industry AI auditing [30]. Second, environmental and epistemic sustainability: training and running large models consumes substantial energy and recursive training on AI-generated material can degrade model quality, raising concerns about the long-term integrity of the scholarly record if synthetic text proliferates in the literature. Both concerns are supported mainly by conceptual analyses and experiments outside research practice (Table 3) [31,90].

Emerging strategies (RQ4)

Ethical standards and guidelines

The first response has been the rapid production of standards. Beyond the publication-ethics guidance in Table 4, research-specific frameworks are emerging [12,91]. Bockting et al. proposed "living guidelines" overseen by an independent scientific body, an approach subsequently reflected in the European Commission's living guidelines on the responsible use of generative AI in research [92]. The broader AI-ethics literature, cited here as context, notes that principles alone lack enforcement mechanisms and may function as ethics-washing [7,93] and emphasises translating principles into tools, processes and accountability structures [8,94,95].

Promoting openness through documentation and reporting

Structured documentation is the most mature practical strategy. Datasheets for datasets record motivation, composition, collection processes and recommended uses while model cards report intended use and disaggregated performance and Data Cards extend this approach to dataset lifecycles [96-98]. Uptake remains partial, however: an analysis of 32,000 model cards found that sections on limitations, bias and environmental impact were frequently missing or minimal [99]. In biomedicine, AI-specific reporting guidelines, namely CONSORT-AI and SPIRIT-AI for trials and TRIPOD+AI for prediction models operationalise transparency for journals and reviewers. Broader open-science infrastructures, including code and data sharing mandates and the Transparency and Openness Promotion (TOP) guidelines provide complementary mechanisms [100-103].

Adapting ethics review and governance

Several authors propose extending the remit of research ethics committees and IRBs to include AI-specific risk assessment, covering dataset provenance, bias and downstream harm, even where research does not involve direct human participants [64-67]. None of the included studies, however, evaluates how such reforms work in practice (Table 3). General-purpose risk-management frameworks, cited here as context, such as the NIST AI RMF, organised around the functions govern, map, measure and manage offer templates that institutions can adapt [53].

Interdisciplinary approaches and capacity building

A recurring recommendation in the corpus is to build AI literacy among researchers, reviewers and ethics-committee members so that they can recognise model limitations, verify outputs and apply disclosure norms consistently [85,104,105]. Embedding ethical reflection within research teams through interdisciplinary collaboration is also widely advocated, consistent with frameworks of responsible research and innovation that emphasise anticipation, reflexivity, inclusion and responsiveness [106].

Discussion

Synthesis: From principles to practice

At the heart of this review is a persistent principle–practice gap and the evidence-gap map shows where it lies. Agreement exists on high-level values and on a limited set of publication rules (no AI authorship, disclosure and human accountability), but these rules, like most of the empirical evidence, apply to the visible end of the research lifecycle: Writing and peer review. The upstream stages of study design and data collection, where the conceptual literature locates the most consequential risks of bias, consent and data governance, are each supported by only four empirical studies and by few stage-specific provisions. Differences in disclosure thresholds, peer-review

https://doi.org/10.70844/ijas.2026.3.57

Innovative Journal of Applied Science

rules and detector use reflect not merely institutional variation but genuine normative uncertainty about where assistance ends and authorship begins.

To help close this gap, Figure 4 presents an integrated framework that treats the five principles of Table 1 as cross-cutting obligations at every stage of the research lifecycle, supported by the governance enablers identified in Section 4.5. Its purpose is to shift ethical attention upstream, towards bias audits and consent review at the data stage and documentation and verification at the analysis stage, rather than relying on disclosure at the publication stage alone.

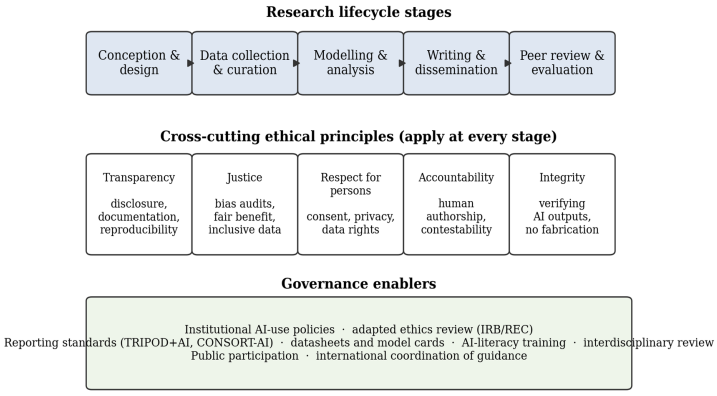

Figure 4: Integrated framework for ethical AI-enabled academic research, linking cross-cutting principles to research lifecycle stages and governance enablers (authors' synthesis).

Contribution relative to existing lifecycle frameworks

Several lifecycle-oriented and research-specific approaches already exist (Table 5). Morley et al. mapped AI-ethics tools against principles and the stages of machine-learning development and the NIST AI Risk Management Framework organises risk management around the functions govern, map, measure and manage [8,53]; neither is specific to research. Research-specific proposals include an analysis of how established research-ethics norms apply to AI the seven-step ETHICAL framework for individual researchers using generative AI [91] and the European Commission's living guidelines addressed to researchers, research organisations and funders [12,91,92]. The framework proposed here consolidates many of the same principles and does not claim novelty at that level. Its distinctive contribution is threefold. First, it is evidence-weighted: each principle–stage combination is linked to the number and type of studies in a mapped corpus (Figure 3, Table 3), so that users can see where recommendations rest on empirical findings and where they rest on argument alone. Second, it identifies a governance–evidence mismatch: both guidance and empirical research concentrate on writing and peer review, whereas the design and data stages, where the conceptual literature locates many of the gravest risks, are the least studied and least governed. Third, it assigns these gaps to actors (Section 5.4), rather than offering principles addressed to individual researchers alone.

Framework	Scope	Organising logic	Evidence base	Gaps and actors
Morley et al. [8]	AI development in general	Principles × stages of machine-learning development	Review of publicly available tools	Identifies tool gaps; not research-specific
NIST [53]	AI systems in general	Functions: govern, map, measure, manage	Multi-stakeholder consensus	Organisational risk processes; not research-specific
Resnik & Hosseini [12]	AI in scientific research	Established research-ethics norms and issues	Conceptual analysis	Calls for new guidance; not evidence-mapped by stage
Eacersall et al. [91]	Generative AI in research	Seven practical principles for researchers (ETHICAL)	Expert, practice-based	Addressed to individual researchers; not stage- or actor-differentiated
European Commission [92]	Generative AI in research	Recommendations by actor (researchers, organisations, funders)	Policy consultation	Actor-specific; not evidence-mapped by stage
This review	AI, including generative AI, in academic research	Five principles × six lifecycle stages	Scoping review of 124 items with evidence-type and confidence ratings	Stage-level evidence and governance gaps, assigned to actors

Table 5: Comparison of the proposed framework with existing lifecycle-oriented and research-specific frameworks.

https://doi.org/10.70844/ijas.2026.3.57

Innovative Journal of Applied Science

Future research directions

Building on the gaps identified, we propose a research agenda organised around four priorities (Table 6).

Direction	Priority research questions	Suggested methods
Upstream empirical research	How is AI used in study design, data collection and data curation and with what ethical consequences for bias, consent and data governance?	Observational and audit studies; ethnography of research teams; analysis of protocols and data-management plans
Context-sensitive ethical frameworks	How should disclosure be graded by type of AI use? When does AI assistance undermine authorship or originality? How can relational justice be operationalised in research design?	Normative analysis; Delphi studies; vignette experiments; framework validation studies
Empirical research on practice	How do researchers actually use and disclose AI? What is the prevalence of undisclosed use and AI-fabricated content?	Large-scale surveys; bibliometric and linguistic analysis; audit studies
Public participation	Which participatory mechanisms improve the legitimacy and equity of AI research? How can affected communities shape dataset governance?	Participatory action research; deliberative forums; comparative case studies
International coordination	How do national and publisher policies diverge and with what consequences? How can Global South perspectives be embedded in standards?	Comparative policy analysis; multi-stakeholder consensus processes
Ethics review and tooling	How effective are adapted REC/IRB processes, datasheets and reporting guidelines at reducing harm?	Pre–post evaluations; randomised trials of reporting interventions; implementation science

Table 6: Proposed agenda for future research.

New ethical frameworks appropriate to AI's complexity. Principist frameworks are troubled by distributed agency, emergent model behavior and the scale of data reuse. Future work should develop and empirically test context-sensitive frameworks, such as relational approaches to algorithmic justice [24] and graded disclosure taxonomies that distinguish editing, analytic and generative uses of AI. Notably, there are few empirical studies examining how researchers actually use AI and how disclosure affects reader trust and reviewer judgment. Above all, empirical studies of how AI is used at the design and data stages, which the evidence-gap map identifies as the least studied, should be a priority.

Enhanced public participation. Research which uses population-scale data and which produces knowledge with societal implications deserves public input into its governance. AI research contexts under-study mechanisms such as citizens' juries, community advisory boards and participatory dataset design. An important future direction is to evaluate their impacts on research legitimacy, data quality and equity, especially for communities that have been historically excluded from or harmed by research.

International collaboration towards global guidelines. There are still major gaps left by the research exemption under the EU AI Act, by the voluntary nature of the UNESCO and OECD instruments and by the Global North focus of the guidance. To counter regulatory arbitrage and ensure guidance reflects diverse research contexts, comparable policy research, harmonization of publisher and funder mandates and the inclusion of Global South institutions in standard-setting are required.

Implications for practice

For researchers, the findings imply a duty to verify every AI-generated claim and reference, to document AI use across the lifecycle rather than only in writing and to avoid uploading confidential or personal data to external tools without authorisation. For institutions, they point to clear, harmonised AI-use policies, AI-competent ethics review that also covers design- and data-stage risks and training. For journals and publishers, they suggest graded disclosure templates, AI-specific reporting guidelines and not relying on detectors alone. For funders and regulators, they highlight the need to close the governance gap left by research exemptions, to fund upstream empirical research and to support international coordination.

For higher education specifically, the findings suggest three priorities. First, research-integrity training for doctoral candidates and staff should include AI literacy: verifying AI-generated references and claims, understanding the limitations of AI-text detectors and documenting AI use across the research lifecycle. Second, ethics-review procedures should extend to design- and data-stage uses of AI, which the evidence-gap map shows are the least studied and least governed. Third, institutional policies should address research uses of AI explicitly and consistently and should not rely on AI-text detectors

as the sole evidence in misconduct cases, given their documented unreliability and bias [41,59].

Limitations

This review has several limitations. First, it is a scoping review: its aim was to map the nature and distribution of the evidence and it does not provide the exhaustive retrieval, risk-of-bias assessment, or certainty-of-evidence grading expected of a systematic review. Second, the search relied on a general scholarly search engine, with retrieval limited to the first page of results for each of 34 strings, rather than on exports from bibliographic databases; despite citation chasing and known-item checks, relevant studies may have been missed and ranking-based results may not be exactly reproducible. Third, screening and charting were performed by a single reviewer and 42% of charting entries are based on titles and abstracts. Fourth, the confidence ratings describe the breadth and type of supporting evidence, not the quality of individual studies. Fifth, 19 records remain awaiting classification, the field is moving fast and the restriction to English-language sources limits generalisability to non-Anglophone and Global South contexts. A full systematic review using database searches, dual independent screening and formal appraisal would be a valuable next step and could build on the search strings and record log provided here.

Summary

AI is revolutionizing the planning, execution, reporting, and assessment of scholarly research. The research community has swiftly agreed upon a minimal set of standards (no AI authorship, disclosure of use, and human accountability), but opinions on disclosure thresholds, peer review, detection, and regulatory scope are still divided, according to this scoping review. The best-supported conclusions relate to the writing and peer-review phases: detectors are biased and unreliable, undisclosed AI use is common, and references are made up. The design and data stages, where risks of bias, consent, and data governance arise, continue to be the least researched and governed, according to the evidence-gap map. Therefore, the evidence-weighted framework presented here assigns the resulting gaps to researchers, institutions, publishers, funders, and regulators while focusing ethical attention upstream. In order to provide globally coherent guidance, future work should give priority to upstream empirical research, ethics-review reform, public participation, and international collaboration. Ethics is a prerequisite for the legitimacy of AI-enabled research, not an external limitation.

Conclusions

AI is changing how academic research is conceptualized, carried out, analysed and published. This scoping review shows that the research community has quickly converged on a minimum set of norms (no AI authorship, disclosure of use, and human accountability), but remains divided on disclosure thresholds, peer review, detection and regulatory scope. The most supported findings are about writing and peer review stages, with pervasive undisclosed use of AI, fake references, and unreliable and biased detectors. The evidence-gap map reveals the design and data stages, the stages where risks of bias, consent, and data governance arise, as the least studied and least governed. So, the evidence-weighted framework proposed here shifts the ethical focus upstream, and leaves the resultant gaps to researchers, institutions, publishers, funders and regulators. Subsequent work needs to focus on upstream empirical research, ethics-review reform, public participation and international collaboration to develop globally coherent guidance. Ethics is not an external constraint on AI-enabled research, but rather a condition of its credibility.

Author contributions

Fredrick M. Ogore conceived and designed the review, conducted the search, screening and data charting, analysed and interpreted the evidence, prepared the figures and wrote and revised the manuscript.

Ethics approval

Not applicable. This review analysed published literature and publicly available documents and did not involve human participants.

Data availability

The search strings, the record-level search and screening log (359 records with decisions and reasons), the data-charting table for the 124 included studies, the evidence-gap map data, the finding-level evidence table.

Conflict of interest

The author declares no conflict of interest.

References

1. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, et al. (2021) Highly accurate protein structure prediction with AlphaFold. Nature 596: 583-589. [Crossref] [GoogleScholar]
2. Wang H, Fu T, Du Y, Gao W, Huang K, et al. (2023) Scientific discovery in the age of artificial intelligence. Nature 620: 47-60. [Crossref] [GoogleScholar]
3. Van Noorden R, Perkel JM (2023) AI and science: What 1,600 researchers think. Nature 621: 672-675. [Crossref] [GoogleScholar]
4. Liang W, Zhang Y, Wu Z, Lepp H, Ji W, et al. (2024) Mapping the increasing use of LLMs in scientific papers. arXiv:2404.01268. [Crossref]
5. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research (1979) The Belmont Report: Ethical principles and guidelines for the protection of human subjects of research. U.S. Department of Health, Education and Welfare.
6. Jobin A, Ienca M, Vayena E (2019) The global landscape of AI ethics guidelines. Nature Machine Intelligence 1: 389-399. [Crossref] [GoogleScholar]

7. Mittelstadt B (2019) Principles alone cannot guarantee ethical AI. Nature Machine Intelligence 1: 501-507. [Crossref] [GoogleScholar]
8. Morley J, Floridi L, Kinsey L, Elhalal A (2020) From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices. Science and Engineering Ethics 26: 2141-2168. [Crossref] [GoogleScholar]
9. Hagendorff T (2020) The ethics of AI ethics: An evaluation of guidelines. Minds and Machines 30: 99-120. [Crossref] [GoogleScholar]
10. Kasneci E, Sessler K, Küchemann S, Bannert M, Dementieva D, et al. (2023) ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences 103: 102274. [Crossref] [GoogleScholar]
11. Lund BD, Wang T, Mannuru NR, Nie B, Shimray S, et al. (2023) ChatGPT and a new academic reality: Artificial intelligence-written research papers and the ethics of the large language models in scholarly publishing. Journal of the Association for Information Science and Technology 74: 570-581. [Crossref] [GoogleScholar]
12. Resnik DB, Hosseini M (2025) The ethics of using artificial intelligence in scientific research: New guidance needed for a new tool. AI and Ethics 5: 1499-1521. [Crossref] [GoogleScholar]
13. Rana NK (2025) Generative AI and academic research: A review of the policies from selected HEIs. Higher Education for the Future 12(1): 97-113. [Crossref] [GoogleScholar]
14. Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, et al. (2018) Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Medical Research Methodology 18: 143. [Crossref] [GoogleScholar]
15. Peters MDJ, Marnie C, Tricco AC, Pollock D, Munn Z, et al. (2020) Updated methodological guidance for the conduct of scoping reviews. JBI Evidence Synthesis 18: 2119-2126. [Crossref] [GoogleScholar]
16. Beauchamp TL, Childress JF (2019) Principles of biomedical ethics. 8th edn. Oxford University Press. [GoogleScholar]
17. Organisation for Economic Co-operation and Development (OECD) (2019) Recommendation of the council on artificial intelligence (OECD/LEGAL/0449). OECD Publishing.
18. UNESCO (2021) Recommendation on the ethics of artificial intelligence. United Nations Educational, Scientific and Cultural Organization.
19. Floridi L, Cowls J (2019) A unified framework of five principles for AI in society. Harvard Data Science Review 1. [Crossref] [GoogleScholar]
20. Barredo Arrieta A, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, et al. (2020) Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion 58: 82-115. [Crossref] [GoogleScholar]
21. Rudin C (2019) Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence 1: 206-215. [Crossref] [GoogleScholar]
22. Haibe-Kains B, Adam GA, Hosny A, Khodakarami F, Massive Analysis Quality Control (MAQC) Society Board of Directors, et al. (2020) Transparency and reproducibility in artificial intelligence. Nature 586: E14-E16. [Crossref] [GoogleScholar]
23. Mehrabi N, Morstatter F, Saxena N, Lerman K, Galstyan A (2021) A survey on bias and fairness in machine learning. ACM Computing Surveys 54(6): 1-35. [Crossref] [GoogleScholar]
24. Birhane A (2021) Algorithmic injustice: A relational ethics approach. Patterns 2: 100205. [Crossref] [GoogleScholar]
25. Crawford K (2021) Atlas of AI: Power, politics and the planetary costs of artificial intelligence. Yale University Press. [Crossref]
26. Metcalf J, Crawford K (2016) Where are human subjects in Big Data research? The emerging ethics divide. Big Data & Society 3. [Crossref] [GoogleScholar]
27. Rocher L, Hendrickx JM, de Montjoye YA (2019) Estimating the success of re-identifications in incomplete datasets using generative models. Nature Communications 10: 3069. [Crossref] [GoogleScholar]
28. Committee on Publication Ethics (COPE) (2023) Authorship and AI tools: COPE position statement.
29. International Committee of Medical Journal Editors (ICMJE) (2024) Recommendations for the conduct, reporting, editing and publication of scholarly work in medical journals.
30. Raji ID, Smart A, White RN, Mitchell M, Gebru T, et al. (2020) Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In Proceedings of the 2020 Conference on Fairness, Accountability and Transparency. ACM 33-44. [Crossref] [GoogleScholar]
31. Strubell E, Ganesh A, McCallum A (2019) Energy and policy considerations for deep learning in NLP. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. ACL 3645-3650. [Crossref] [GoogleScholar]
32. Bender EM, Gebru T, McMillan-Major A, Shmitchell S (2021) On the dangers of stochastic parrots: Can language models be too big? In Proceedings of the 2021 ACM Conference on Fairness, Accountability and Transparency. ACM 610-623. [Crossref] [GoogleScholar]
33. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, et al. (2018) PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. Annals of Internal Medicine 169: 467-473. [Crossref] [GoogleScholar]
34. Arksey H, O'Malley L (2005) Scoping studies: Towards a methodological framework. International Journal of Social Research Methodology 8: 19-32. [Crossref] [GoogleScholar]
35. Levac D, Colquhoun H, O'Brien KK (2010) Scoping studies: Advancing the methodology. Implementation Science 5: 69. [Crossref] [GoogleScholar]
36. Snilstveit B, Vojtkova M, Bhavsar A, Stevenson J, Gaarder M (2016) Evidence & gap maps: A tool for promoting evidence

informed policy and strategic research agendas. Journal of Clinical Epidemiology 79: 120-129. [Crossref] [GoogleScholar]

37. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, et al. (2021) The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ 372: n71. [Crossref] [GoogleScholar]
38. Gray A (2024) ChatGPT "contamination": Estimating the prevalence of LLMs in the scholarly literature. arXiv:2403.16887. [Crossref] [GoogleScholar]
39. Kobak D, González-Márquez R, Horvát EÁ, Lause J (2025) Delving into LLM-assisted writing in biomedical publications through excess vocabulary. Science Advances 11: eadt3813. [Crossref] [GoogleScholar]
40. Walters WH, Wilder EI (2023) Fabrication and errors in the bibliographic citations generated by ChatGPT. Scientific Reports 13: 14045. [Crossref] [GoogleScholar]
41. Weber-Wulff D, Anohina-Naumeca A, Bjelobaba S, Foltýnek T, Guerrero-Dib J, et al. (2023) Testing of detection tools for AI-generated text. International Journal for Educational Integrity 19: 26. [Crossref] [GoogleScholar]
42. Hartung J, Reuter S, Kulow VA, Fähling M, Spreckelsen C, et al. (2024) Experts fail to reliably detect AI-generated histological data. Scientific Reports 14. [Crossref] [GoogleScholar]
43. Ganjavi C, Eppler MB, Pekcan A, Biedermann B, Abreu A, et al. (2024) Publishers' and journals' instructions to authors on use of generative artificial intelligence in academic and scientific publishing: Bibliometric analysis. BMJ 384: e077192. [Crossref] [GoogleScholar]
44. Lund BD, Naheem KT (2024) Can ChatGPT be an author? A study of artificial intelligence authorship policies in top academic journals. Learned Publishing 37: 13-21. [Crossref] [GoogleScholar]
45. Stokel-Walker C (2023) ChatGPT listed as author on research papers: Many scientists disapprove. Nature 613: 620-621. [Crossref] [GoogleScholar]
46. Thorp HH (2023) ChatGPT is fun, but not an author. Science 379: 313. [Crossref] [GoogleScholar]
47. Zielinski C, Winker MA, Aggarwal R, Ferris LE, Heinemann M, et al. (2023) Chatbots, generative AI and scholarly manuscripts: WAME recommendations on chatbots and generative artificial intelligence in relation to scholarly publications. World Association of Medical Editors.
48. Flanagin A, Kendall-Taylor J, Bibbins-Domingo K (2023) Guidance for authors, peer reviewers and editors on use of AI, language models and chatbots. JAMA 330: 702-703. [Crossref] [GoogleScholar]
49. Kaebnick GE, Magnus DC, Kao A, Hosseini M, Resnik D, et al. (2023) Editors' statement on the responsible use of generative AI technologies in scholarly journal publishing. Hastings Center Report 53: 3-6. [Crossref] [GoogleScholar]
50. Hosseini M, Resnik DB, Holmes K (2023) The ethics of disclosing the use of artificial intelligence tools in writing scholarly manuscripts. Research Ethics 19: 449-465. [Crossref] [GoogleScholar]
51. Nature (2023) Tools such as ChatGPT threaten transparent science; here are our ground rules for their use. Nature 613: 612. [Crossref]
52. Elsevier (2023) The use of generative AI and AI-assisted technologies in writing for Elsevier.
53. National Institute of Standards and Technology (NIST) (2023) Artificial intelligence risk management framework (AI RMF 1.0) (NIST AI 100-1). [Crossref]
54. Hosseini M, Horbach SPJM (2023) Fighting reviewer fatigue or amplifying bias? Considerations and recommendations for use of ChatGPT and other large language models in scholarly peer review. Research Integrity and Peer Review 8: 4. [Crossref] [GoogleScholar]
55. Liang W, Izzo Z, Zhang Y, Lepp H, Cao H, et al. (2024) Monitoring AI-modified content at scale: A case study on the impact of ChatGPT on AI conference peer reviews. In Proceedings of the 41st International Conference on Machine Learning (PMLR 235). [Crossref] [GoogleScholar]
56. Latona GR, Ribeiro MH, Davidson TR, Veselovsky V, West R (2025) The AI review lottery: Widespread AI-assisted peer reviews boost paper scores and acceptance rates. Proceedings of the ACM on Human-Computer Interaction. [Crossref] [GoogleScholar]
57. Mollaki V (2024) Death of a reviewer or death of peer review integrity? The challenges of using AI tools in peer reviewing and the need to go beyond publishing policies. Research Ethics 20: 239-250. [Crossref] [GoogleScholar]
58. Elkhatat AM, Elsaid K, Almeer S (2023) Evaluating the efficacy of AI content detection tools in differentiating between human and AI-generated text. International Journal for Educational Integrity 19: 17. [Crossref] [GoogleScholar]
59. Liang W, Yuksekgonul M, Mao Y, Wu E, Zou J (2023) GPT detectors are biased against non-native English writers. Patterns 4: 100779. [Crossref] [GoogleScholar]
60. European Union (2024) Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union, L series.
61. Resnik DB, Hosseini M (2025) Disclosing artificial intelligence use in scientific research and publication: When should disclosure be mandatory, optional, or unnecessary? Accountability in Research 33. [Crossref] [GoogleScholar]
62. Hosseini M, Gordijn B, Kaebnick GE, Holmes K (2025) Disclosing generative AI use for writing assistance should be voluntary. Research Ethics 21(4): 728-735. [Crossref] [GoogleScholar]
63. Elsevier (2024) Insights 2024: Attitudes toward AI. Elsevier.
64. Friesen P, Douglas-Jones R, Marks M, Pierce R, Fletcher K, et al. (2021) Governing AI-driven health research: Are IRBs up to the task? Ethics & Human Research 43: 35-42. [Crossref] [GoogleScholar]

65. Ferretti A, Ienca M, Sheehan M, Blasimme A, Dove ES, et al. (2021) Ethics review of big data research: What should stay and what should be reformed? BMC Medical Ethics 22: 51. [Crossref] [GoogleScholar]
66. Bouhouita-Guermech S, Gogognon P, Bélisle-Pipon JC (2023) Specific challenges posed by artificial intelligence in research ethics. Frontiers in Artificial Intelligence 6: 1149082. [Crossref] [GoogleScholar]
67. Secretary's Advisory Committee on Human Research Protections (SACHRP) (2022) Considerations for IRB review of research involving artificial intelligence. U.S. Department of Health and Human Services.
68. Nanayakkara P, Hullman J, Diakopoulos N (2021) Unpacking the expressed consequences of AI research in broader impact statements. In Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics and Society. ACM 795 - 806. [Crossref] [GoogleScholar]
69. Ashurst C, Hine E, Sedille P, Carlier A (2022) AI ethics statements: Analysis and lessons learnt from NeurIPS broader impact statements. In Proceedings of the 2022 ACM Conference on Fairness, Accountability and Transparency. ACM 2047-2056. [Crossref] [GoogleScholar]
70. Liu D, Nanayakkara P, Sakha SA, Abuhamad G, Blodgett SL, et al. (2022) Examining responsibility and deliberation in AI impact statements and ethics reviews. In Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics and Society. ACM 424-435. [Crossref] [GoogleScholar]
71. Birhane A, Kalluri P, Card D, Agnew W, Dotan R, et al. (2022) The values encoded in machine learning research. In Proceedings of the 2022 ACM Conference on Fairness, Accountability and Transparency. ACM 173-184. [Crossref] [GoogleScholar]
72. Dillion D, Tandon N, Gu Y, Gray K (2023) Can AI language models replace human participants? Trends in Cognitive Sciences 27: 597-600. [Crossref] [GoogleScholar]
73. Grossmann I, Feinberg M, Parker DC, Christakis NA, Tetlock PE, et al. (2023) AI and the transformation of social science research. Science 380: 1108-1109. [Crossref] [GoogleScholar]
74. Bail CA (2024) Can generative AI improve social science? Proceedings of the National Academy of Sciences 121: e2314021121. [Crossref] [GoogleScholar]
75. Buolamwini J, Gebru T (2018) Gender shades: Intersectional accuracy disparities in commercial gender classification. Proceedings of Machine Learning Research 81: 77-91. [GoogleScholar]
76. Obermeyer Z, Powers B, Vogeli C, Mullainathan S (2019) Dissecting racial bias in an algorithm used to manage the health of populations. Science 366: 447-453. [Crossref] [GoogleScholar]
77. Selbst AD, Boyd D, Friedler SA, Venkatasubramanian S, Vertesi J (2019) Fairness and abstraction in sociotechnical systems. In Proceedings of the Conference on Fairness, Accountability and Transparency. ACM, pp. 59–68. [Crossref] [GoogleScholar]
78. Zook M, Barocas S, Boyd D, Crawford K, Keller E, et al. (2017) Ten simple rules for responsible big data research. PLOS Computational Biology 13: e1005399. [Crossref] [GoogleScholar]
79. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3: 160018. [Crossref] [GoogleScholar]
80. Alkaissi H, McFarlane SI (2023) Artificial hallucinations in ChatGPT: Implications in scientific writing. Cureus 15: e35179. [Crossref] [GoogleScholar]
81. Bhattacharyya M, Miller VM, Bhattacharyya D, Miller LE (2023) High rates of fabricated and inaccurate references in ChatGPT-generated medical content. Cureus 15: e39238. [Crossref] [GoogleScholar]
82. Chelli M, Descamps J, Lavoué V, Trojani C, Azar M, et al. (2024) Hallucination rates and reference accuracy of ChatGPT and Bard for systematic reviews: Comparative analysis. Journal of Medical Internet Research 26: e53164. [Crossref] [GoogleScholar]
83. Kendall G, Teixeira da Silva JA (2024) Risks of abuse of large language models, like ChatGPT, in scientific publishing: Authorship, predatory publishing and paper mills. Learned Publishing 37: 55-62. [Crossref] [GoogleScholar]
84. Gu J, Wang X, Li C, Zhao J, Fu W, et al. (2022) AI-enabled image fraud in scientific publications. Patterns 3: 100511. [Crossref] [GoogleScholar]
85. Binz M, Alaniz S, Roskies A, Aczel B, Bergstrom CT, et al. (2025) How should the advancement of large language models affect the practice of science? Proceedings of the National Academy of Sciences 122: e2401227121. [Crossref] [GoogleScholar]
86. Gundersen OE, Kjensmo S (2018) State of the art: Reproducibility in artificial intelligence. Proceedings of the AAAI Conference on Artificial Intelligence 32: 1644-1651. [Crossref] [GoogleScholar]
87. Kapoor S, Narayanan A (2023) Leakage and the reproducibility crisis in machine-learning-based science. Patterns 4: 100804. [Crossref] [GoogleScholar]
88. Gao CA, Howard FM, Markov NS, Dyer EC, Ramesh S, et al. (2023) Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. npj Digital Medicine 6: 75. [Crossref] [GoogleScholar]
89. Messeri L, Crockett MJ (2024) Artificial intelligence and illusions of understanding in scientific research. Nature 627: 49-58. [Crossref] [GoogleScholar]
90. Shumailov I, Shumaylov Z, Zhao Y, Papernot N anderson R, et al. (2024) AI models collapse when trained on recursively generated data. Nature 631: 755–759. [Crossref] [GoogleScholar]
91. Eacersall D, Pretorius L, Smirnov I, Spray E, Illingworth S, et al. (2024) Navigating ethical challenges in generative AI-enhanced research: The ETHICAL framework for responsible

generative AI use. arXiv:2501.09021. [Crossref] [GoogleScholar]

92. European Commission (2024) Living guidelines on the responsible use of generative AI in research. Directorate-General for Research and Innovation, ERA Forum.
93. Munn L (2023) The uselessness of AI ethics. AI and Ethics 3: 869-877. [Crossref] [GoogleScholar]
94. Whittlestone J, Nyrup R, Alexandrova A, Cave S (2019) The role and limits of principles in AI ethics: Towards a focus on tensions. In Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics and Society. ACM 195-200. [Crossref] [GoogleScholar]
95. Bockting CL, van Dis EAM, van Rooij R, Zuidema W, Bollen J (2023) Living guidelines for generative AI — why scientists must oversee its use. Nature 622: 693-696. [Crossref] [GoogleScholar]
96. Gebru T, Morgenstern J, Vecchione B, Vaughan JW, Wallach H, et al. (2021) Datasheets for datasets. Communications of the ACM 64: 86-92. [Crossref] [GoogleScholar]
97. Mitchell M, Wu S, Zaldivar A, Barnes P, Vasserman L, et al. (2019) Model cards for model reporting. In Proceedings of the Conference on Fairness, Accountability and Transparency. ACM 220-229. [Crossref] [GoogleScholar]
98. Pushkarna M, Zaldivar A, Kjartansson O (2022) Data cards: Purposeful and transparent dataset documentation for responsible AI. In Proceedings of the 2022 ACM Conference on Fairness, Accountability and Transparency. ACM 1776-1826. [Crossref] [GoogleScholar]
99. Liang W, Rajani N, Yang X, Ozoani E, Wu E, et al. (2024) What's documented in AI? Systematic analysis of 32K AI model cards. arXiv:2402.05160. [Crossref] [GoogleScholar]
100. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK, et al. (2020) Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: The CONSORT-AI extension. Nature Medicine 26: 1364-1374. [Crossref] [GoogleScholar]
101. Cruz Rivera S, Liu X, Chan AW, Denniston AK, Calvert MJ, et al. (2020) Guidelines for clinical trial protocols for interventions involving artificial intelligence: The SPIRIT-AI extension. Nature Medicine 26: 1351-1363. [Crossref] [GoogleScholar]
102. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, et al. (2024) TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ 385: e078378. [Crossref] [GoogleScholar]
103. Nosek BA, Alter G, Banks GC, Borsboom D, Bowman SD, et al. (2015) Promoting an open research culture. Science 348: 1422-1425. [Crossref] [GoogleScholar]
104. Dwivedi YK, Kshetri N, Hughes L, Slade EL, Jeyaraj A, et al. (2023) "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. International Journal of Information Management 71: 102642. [Crossref] [GoogleScholar]
105. van Dis EAM, Bollen J, Zuidema W, van Rooij R, Bockting CL (2023) ChatGPT: Five priorities for research. Nature 614: 224-226. [Crossref] [GoogleScholar]
106. Stilgoe J, Owen R, Macnaghten P (2013) Developing a framework for responsible innovation. Research Policy 42: 1568-1580. [Crossref] [GoogleScholar].

Copyright: © 2026 The Author(s). Published by Innovative Journal of Applied Science. This is an open-access article under the terms of the Creative Commons Attribution License (CC BY). (https://creativecommons.org/licenses/by/4.0/).