Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Governance Framework for Human–AI Collaboration in Academic Writing: The Scholarly HI–AI Loop

Thamburaj Anthuvan
USV (India)
Lakshmi Nagendra
JSS Academy of Higher Education & Research
Sunitha Prabhuram
sunitha.prabhuram@manipaldubai.com

Manipal Academy of Higher Education (MAHE), Dubai, UAE
Om J. Lakhani
AB Plus hospital
Sanjay Kalra
Bharti Hospital
Kajal Maheshwari
PCET's S.B. Patil Institute of Management
Sandeep Srivathsan
Great Lakes Institute of Management
Sailatha Ravi
PGx India

Research Article

Keywords: Human AI collaboration, AI governance, academic writing, research integrity, literature synthesis, responsible AI

Posted Date: September 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10644052/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Human– artificial intelligence (AI) collaboration is transforming academic writing by enabling rapid literature screening, summarization, and draft generation. Despite these benefits, concerns remain regarding factual accuracy, contextual interpretation, transparency, and research integrity, reflecting inconsistent governance and uneven institutional guidance. This study develops an evidence-informed governance framework, the Scholarly Human Intelligence–Artificial Intelligence (HI–AI) Loop, to structure human oversight in AI-assisted academic writing. A structured narrative review of peer-reviewed literature published between 2020 and 2025 was conducted to examine AI- assisted scholarly workflows, the extent of human involvement, and reported ethical and institutional governance practices. Of the 87 records identified through database searches, 38 studies met the eligibility criteria. The reviewed literature indicated that AI was most frequently used for literature screening, summarization, and draft preparation, with reported reductions in time and effort required. Common limitations included citation inaccuracies, limited contextual interpretation, and inconsistent disclosure of AI use, with such disclosure absent in 12 of the 38 included studies. More than half of the studies described hybrid workflows integrating AI assistance with structured human oversight, whereas fewer reported AI-only approaches. Across the reviewed literature, structured human oversight is generally associated with improved factual accuracy, interpretive coherence, and ethical accountability. Based on this synthesis, the proposed seven-stage Scholarly HI AI Loop offers a governance-oriented framework for systematically integrating human judgment into AI-assisted scholarly workflows, with potential applicability across disciplines and institutional settings.

1. Introduction

The expansion of scholarly publishing has increased the demand for academic writing and literature synthesis, particularly in disciplines where evidence synthesis informs clinical practice, policy, and research decision-making [1], [2]. Systematic reviews and narrative syntheses remain central to evidence-based research; however, the growing volume and complexity of the literature are making traditional methods less practical, requiring considerable time and resources [3], [4]. In response, researchers and institutions have adopted computer-assisted tools to support literature retrieval, summarization, and early stage drafting [5], [6]. Advances in language-based artificial intelligence (AI) systems have enabled faster abstract screening, thematic clustering, and text generation, thereby supporting the handling of large volumes of information [7]. However, evidence indicates that these systems frequently generate outputs with limited contextual understanding and variable citation accuracy, raising concerns regarding their reliability and scholarly integrity [8]. Improvements in linguistic fluency have not addressed the underlying limitations of interpretive reasoning, disciplinary nuance, and ethical judgment [9].

Consequently, AI-only approaches remain insufficient for tasks that require evidence appraisal, interpretive judgment, and accountability. Recent studies suggest that hybrid models, which combine AI assistance with human oversight, are becoming more common, positioning AI as a supportive tool rather

than an autonomous academic entity [10], [11]. Despite this shift, the editorial, ethical, and institutional implications of hybrid workflows remain poorly defined [12], [13]. The use of AI in academic writing is further complicated by unresolved questions concerning authorship attribution, plagiarism prevention, intellectual property rights, and equitable access [14]–[17]. Although publishers and professional organizations have issued guidance on AI use, current recommendations remain inconsistent and lack a cohesive operational framework for embedding human judgment across different stages of academic writing.

Accordingly, there remains an unmet need for a structured approach that explicitly defines how human involvement and responsibility should be embedded in AI-assisted academic writing and literature synthesis. This study addresses this gap by synthesizing existing evidence to propose the Scholarly Human Intelligence–Artificial Intelligence (HI–AI) Loop, a governance-oriented framework designed to guide the systematic integration of human judgment into AI-assisted scholarly workflows.

Established perspectives on human–AI collaboration, including hybrid intelligence and human-in-the-loop models, provide a useful context for understanding how cognitive responsibility and accountability are shared between humans and AI systems [18]–[21]. Sociotechnical perspectives further show how AI-assisted writing is shaped by the institutional norms, editorial practices and governance structures of scholarly publishing [22]. Although the Scholarly HI–AI Loop proposed here is derived from a synthesis of the literature rather than from a single theory, these perspectives offer a conceptual backdrop that reinforces the ethical case for structured human oversight in AI-assisted academic writing.

2. Materials and Methods

This study adopted a narrative review design to develop the Scholarly HI–AI Loop, a conceptual governance framework for human–AI collaboration in academic writing and literature summarization tasks. The review followed structured procedures to ensure transparency and reproducibility, including predefined eligibility criteria, multistage screening, dual-reviewer assessment, and systematic record tracking. The study selection process is summarized in a Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 flow diagram (Fig. 1) [23], [24].

2.1. Database and Search Strategy

A multi-database search was conducted across Scopus, Web of Science, PubMed, and Google Scholar to capture literature spanning biomedical, technological, and social science perspectives on human–AI collaboration in academic writing literature. The search terms combined three concept domains:

("AI-assisted summarization" OR "machine-assisted literature review" OR "large language model summarization" OR "large language models")

AND

("human-in-the-loop" OR "editorial oversight" OR "hybrid workflow")

AND

("scholarly publishing" OR "academic writing")

The search strings were adapted to the indexing syntax of each database. In addition to standard database searches, the Consensus AI discovery engine was used to identify additional peer-reviewed studies using semantic relevance and citation-aware ranking to guide the selection process. Additional search terms included human-in-the-loop summarization, AI-assisted literature review, large language models (LLMs) collaboration in academic writing, and hybrid human–machine knowledge production. Studies identified through this process were screened using the same eligibility criteria as the database-derived records [4], [12], [21].

2.2. Time Frame

Studies published between January 2020 and December 2025 were included. This period captures the rapid emergence and widespread adoption of LLM in academic workflows and corresponds to the most active phase of research on human–AI collaboration in scholarly writing and literature synthesis [3], [6].

2.3. Inclusion and Exclusion Criteria

Eligible studies included peer-reviewed journal articles, conference papers, editorials, and case studies that examined human–AI collaboration in academic writing, literature synthesis or related scholarly tasks. Studies were included if they explicitly described human oversight activities such as fact-checking, critical appraisal, editorial decision-making, ethical review, or interpretive validation. Only English-language full-text publications providing sufficient methodological or conceptual details were eligible. Studies focusing exclusively on technical model development without human involvement or real-world scholarly application, non-peer-reviewed publications, and articles lacking adequate methodological or ethical transparency were excluded from the review.

2.4. Screening Process

Records identified through database searches and supplementary sources underwent a two-stage screening process comprising title and abstract screening, followed by a full-text eligibility assessment. Screening was conducted independently by two reviewers, with disagreements resolved through discussion and, where necessary, adjudication by a third reviewer. The study selection process is summarized in Fig. 1, which is adapted from the PRISMA 2020 framework [23], [24].

2.5. Data Extraction

Data were extracted using a structured data charting form to ensure consistency across studies. The extracted variables included bibliographic characteristics, AI tool type, workflow classification (AI-only or hybrid human–AI), human oversight activities, disclosure practices, reported outcomes, limitations, and

governance-related challenges associated with AI-assisted academic writing. Data extraction was performed independently by two reviewers, and discrepancies were resolved through consensus.

2.6. Quality Control

Independent dual-reviewer screening and data extraction, together with third-reviewer adjudication of disagreements, enhanced methodological consistency and minimized selection bias across study selection, data extraction and thematic classification.

2.7. AI Use Transparency

Consensus AI was used during the literature search to assist in identifying potentially relevant publications. Generative artificial intelligence (GenAI) tools were not used for study selection, data extraction, evidence interpretation, framework development, or autonomous content generation. GenAI-assisted editing was limited to improving grammar, readability, and presentation of author-generated text. All conceptual development, literature synthesis, interpretation, and final approval were undertaken by the authors, who accept full responsibility for the content of this manuscript.

3. Results

A total of 38 peer-reviewed studies were included in the narrative synthesis, examining the use of AI tools in academic writing and literature summarization across biomedical sciences, public health, information science, and related fields. Five consistent thematic patterns emerged across the reviewed literature: (i) epistemic friction and early editorial responses to AI-assisted writing, (ii) the roles and capabilities of AI tools, (iii) the extent and forms of human editorial oversight, (iv) ethical guidelines and disclosure practices, and (v) institutional readiness and barriers to adoption. Each theme is discussed below.

3.1. Narrative Synthesis

3.1.1. Epistemic Friction and Editorial Pushback (2020–2023)

Studies published between 2020 and 2023 consistently report tension between the expanding use of AI tools and concerns about epistemic reliability. Early applications of models such as Generative Pre-trained Transformer 2 (GPT-2) and GPT-3 focused on outlining, paraphrasing, and summarizing literature. Several studies have reported improvements in writing speed and surface coherence but have also identified weak source grounding and limited evidentiary verification [3]. Following the public release of ChatGPT in late 2022, its use has increased rapidly across different fields [7], [25]. Simultaneously, several studies have documented recurring problems, such as hallucinated citations, fake references, and failure to distinguish credible from non-credible sources [4], [7], [25]. These issues have prompted scrutiny from journals, universities, and funding organizations.

By 2023, the editorial responses had become more explicit. Major publishers have issued policies prohibiting the attribution of authorship to large language models or restricting the use of AI-generated content [14], [15], [26]. Scholarly reactions were mixed: some studies supported these restrictions as necessary for accountability, while others questioned their consistency, given existing editorial outsourcing practices [16]. Guidance issued during this period remained fragmented and reactive [17], [16], [14]. Collectively, the literature reflects a shift from exploratory use toward governance-focused discussions emphasizing structured human oversight [19], [22], [27].

3.1.2. AI Tool Roles and Capabilities

Across the reviewed studies, AI tools in academic summarization have evolved from extractive to abstractive approaches. Early extractive systems primarily selected verbatim text segments, offering efficiency but limited thematic integration and contextual interpretation [1]. The introduction of large language models, such as GPT-4 and ChatGPT, has enabled abstractive summarization with improved narrative flow and linguistic coherence [5], [10], [25]. However, multiple studies continue to report factual inaccuracies, hallucinated citations, and poor alignment with methodological standards, particularly in evidence-based domains [4], [11].

Several studies have described efforts to address these limitations using domain adaptation and hybrid architectures. Fine-tuning models on discipline-specific corpora improved adherence to reporting standards, such as PRISMA [6], [3]. However, concerns regarding model opacity, data provenance, and reproducibility persist, particularly for proprietary systems [22], [27], [28]. Despite these concerns, retrieval-augmented workflows that combine LLMs with structured databases have demonstrated improved factual consistency in pharmacotherapy reviews, library science, and public health applications [8], [20], [29]. Across studies, interpretive judgment and contextual framing remained dependent on human input.

3.1.3. Human Editorial Oversight

Human involvement has been consistently described as integral to AI-assisted academic writing rather than a post-processing step [4], [12], [29]. Across studies, human contributors performed fact-checking, narrative restructuring, bias identification, and contextual alignment to ensure disciplinary accuracy and ethical compliance [4], [11]–[12]. One pharmacotherapy review documented the iterative refinement of AI-generated drafts by clinical experts to align the content with therapeutic standards and domain-specific terminology [29]. Other studies have characterized similar practices as “intellectual interventions,” emphasizing the role of human judgment in shaping scholarly coherence and credibility [12].

In parallel, several publishers, including Elsevier, Springer, and SAGE, have introduced policies requiring the disclosure of AI use and clarification of human oversight [30]. These practices align with governance models such as *Artificial Intelligence–Conscious to Conscience (AI-C2C)*, which positions human validation as an embedded checkpoint within hybrid knowledge workflows [27]. Equity-focused critiques further emphasize the need for training and transparent oversight to prevent the concentration of

responsible AI use within well-resourced institutions [31]. The core editorial roles identified across the studies are summarized in Table 1.

Table 1
Human editorial roles in AI-assisted academic writing across reviewed studies

Editorial Role	Description	Representative Studies
Content Curation	Selecting, organizing, and prioritizing AI-generated outputs for relevance and inclusion.	[3]
Fact-Checking & Validation	Verifying the accuracy of AI-produced content, correcting errors, and cross-referencing with source materials.	[4]
Narrative Restructuring	Refining AI drafts to enhance readability, logical flow, and disciplinary framing.	[12], [29]
Bias Identification & Mitigation	Detecting and addressing cultural, disciplinary, or systemic biases in AI-generated content.	[11], [28].
Ethical & Editorial Oversight	Ensuring compliance with ethical frameworks, publisher policies, and scholarly integrity standards.	[27], [30]
Contextualization	Embedding AI-generated content within appropriate theoretical, cultural, or clinical contexts.	[22], [32]

Note. Roles synthesized from recurring practices across the 38 included studies; "editorial oversight" encompasses ethical review and compliance with publisher policies.

3.1.4. Ethical and Disclosure Practices

The ethical integration of AI tools is closely linked to transparency and attribution across the reviewed literature [30], [33]. Among the 38 included studies, 12 did not disclose AI use, despite a clear reliance on generative tools [6], [20]. In studies that reported AI involvement, the level of detail varied substantially. Several studies disclosed only the tool name without specifying whether AI contributed to drafting, synthesis, citation management or editorial review [15]. This lack of reporting granularity limits reproducibility and complicates peer review [34]. Beyond attribution, studies have reported risks related to hallucinated citations, implicit bias, and the reinforcement of dominant epistemic frames, particularly in biomedicine and policy research [3], [16], [29]. Governance approaches, such as AI-C2C, emphasize structured human validation to mitigate these risks [27]. In response, some journals have introduced structured disclosure templates requiring the specification of AI tools, functions, and oversight mechanisms [15], [30], [33]. The identified disclosure practices are summarized in Table 2.

Table 2
Disclosure and ethical practices in AI-assisted academic writing across reviewed studies

Practice	Description	Representative Sources
AI Use Disclosure	Explicitly stating the AI tool used and its role in the review process.	[30], [33]
Granularity of Reporting	Detailing whether AI contributions were limited to drafting or extended to synthesis, restructuring, or citation management.	[15]
Structured Templates	Use of journal-mandated disclosure forms specifying tools, tasks, and oversight mechanisms.	[30], [33]
Ethical Oversight	Incorporating deliberate human validation of AI outputs for accuracy, context, and ethical alignment.	[27], [29]
Bias & Hallucination Mitigation	Human review to detect and correct AI-generated inaccuracies, hallucinated references, and cultural or disciplinary biases.	[16], [28].

Note: *Practices drawn from publisher guidance and reviewed studies.*

3.1.5. Institutional Readiness and Barriers

Several studies have reported that AI-assisted summarization reduces manual workload and supports higher-order scholarly tasks, particularly in healthcare, education, and policy research [7], [21], [29]. Hybrid workflows enable the redistribution of effort toward contextual interpretation and ethical validation [1], [3], [6]. Early-career researchers have noted pedagogical benefits, such as enhanced comprehension of source integration and narrative construction facilitated by supervised AI applications [4]. Potential benefits were also noted for low- and middle-income countries, where hybrid workflows may reduce labor barriers and expand participation in global scholarship [12].

However, institutional adoption remains uneven. Many organizations lack formal protocols for AI disclosure and oversight, raising concerns about transparency and reproducibility [16], [30]. Technical limitations persisted, with instances of hallucination and domain misrepresentation reported even in fine-tuned models [25], [29]. While governance frameworks such as AI-C2C offer mechanisms for embedding human accountability [27], their implementation remains limited, particularly in settings affected by infrastructural and training constraints [12]. The comparative characteristics of the traditional, AI-only, and hybrid workflows are summarized in Table 3.

Table 3
Comparison of traditional, AI-only, and hybrid approaches in academic writing workflows

Criteria	Traditional	AI-Assisted (Minimal Human Oversight)	Hybrid (Human + AI)
Speed	Low (weeks–months)	High (minutes–hours)	Moderate (hours–days)
Accuracy	High (expert-reviewed)	Variable (model-dependent)	High (expert-validated)
Contextual Understanding	Strong	Moderate	Strong (human-augmented)
Citation Integrity	Reliable	Variable; requires verification	Reliable (post-verification)
Human Effort Required	High	Low	Moderate
Risk of Hallucination	Low	High	Low–moderate
Ethical Oversight	Explicit and central	Variable; often limited	Integrated throughout
Scalability	Limited	High	High
Suitability for Niche Domains	High	Limited	High

Note. *Criteria adapted from* [3] and [6]. [1] *and* [27].

3.2. Conceptual Contribution: The Scholarly HI-AI Loop

Based on the patterns identified in the narrative synthesis, this study proposes the Scholarly HI–AI Loop, a conceptual framework for structuring hybrid human–AI collaboration in academic literature summarization. The framework responds to three recurring needs reported across the reviewed studies: efficiency, achieved through AI-assisted screening and drafting [3], [5], scholarly rigor, achieved through human-led validation and contextualization [29], and ethical accountability, achieved through governance mechanisms such as AI-C2C [22], [27]. The HI–AI Loop is organized into seven interdependent stages that formalize points of human involvement across scoping, screening, drafting, refinement, contextualization, ethical validation, and publication. The stages and associated human and AI roles are summarized in Table 4, and the bidirectional collaboration between the contributors and AI tools is illustrated in Fig. 2.

Table 4
Stages of the Scholarly HI-AI Loop for literature summarization

Stage	Purpose	Human Roles	AI Roles	Expected Outputs
Stage 1: Scoping and Selection	Define review objectives, inclusion/exclusion criteria, and search strategy.	Authors and editors define the scope, refine search queries, and ensure disciplinary relevance.	Expand search queries, cluster keywords, and identify potentially relevant sources.	Defined review protocol and search strategy.
Stage 2: Screening and Shortlisting	Identify studies for detailed review.	Reviewers validate AI-generated shortlists and assess borderline cases.	Deduplicate records, screen titles and abstracts, and cluster topics.	Curated set of eligible studies.
Stage 3: Draft Summarization	Generate preliminary evidence summaries.	Review outputs, identify inaccuracies, and assess coherence.	Produce structured abstractive or extractive summaries.	Initial AI-generated summaries.
Stage 4: Human-in-the-Loop Refinement	Improve factual accuracy and contextual depth.	Verify facts, strengthen interpretation, restructure narratives, and incorporate disciplinary expertise.	Support iterative revision and language refinement.	Contextually enriched and validated summaries.
Stage 5: Editorial Contextualization	Align content with scholarly and disciplinary standards.	Integrate theoretical perspectives, harmonize terminology, and strengthen interpretation.	Support section organization and stylistic consistency.	Discipline-aligned scholarly manuscript.
Stage 6: Ethical Validation and Disclosure	Ensure transparency, integrity, and responsible AI use.	Verify citations, assess originality, and prepare AI-use disclosures.	Assist citation verification and generate disclosure records.	Ethically validated manuscript.
Stage 7: Polishing and Publication	Prepare the manuscript for submission.	Proofread, ensure journal compliance, and complete final quality checks.	Improve grammar, formatting, and language consistency.	Submission-ready manuscript.

Note: *Author-developed framework informed by hybrid intelligence theory, human-in-the-loop models, and AI-enhanced collaborative theory.*

Figure 2. visualizes the Scholarly HI–AI Loop, illustrating the bidirectional, co-regulated collaboration between human contributors and AI tools across the seven summarization stages, as detailed in Table 4.

By embedding structured human participation throughout the workflow, particularly during contextualization, ethical review, and disclosure, the Scholarly HI–AI Loop addresses the epistemic, editorial, and institutional gaps identified in the reviewed literature [4], [16]. The framework positions scholars as active agents who oversee AI-assisted processes rather than as passive validators of machine-generated output, thereby reinforcing human accountability, interpretive judgment, and responsible governance throughout the scholarly writing process [17], [26].

4. Discussion

4.1. Framing the Future: What This Study Adds

This study contributes a governance-oriented perspective to the debates on AI use in academic writing by shifting attention from tool performance to the epistemic and institutional conditions under which human–AI collaboration occurs. The literature increasingly suggests that governance has become as important as technological capability. [14], [19], [27]. The Scholarly HI–AI Loop introduced in this review provides a seven-stage framework for structuring human–AI collaboration across the academic writing life cycle. Rather than positioning AI as a shortcut or substitute for scholarly labor, the framework situates it within a guided workflow that depends on transparency and active human oversight from scoping to disclosure. This approach aligns with hybrid intelligence perspectives that emphasize shared cognition and feedback between human and machine actors [18], [35], as well as recent formulations of AI-enhanced collaborative theory.

Beyond conceptual categorization, the HI–AI Loop functions as a governance structure for implementing AI-integrated academic practices [22]. It addresses a recurring gap identified in the literature: the absence of interdisciplinary, multi-actor models that operationalize human-in-the-loop scholarship in the context of GenAI [19], [27]. This offers a practical reference point for researchers, editors, and institutions navigating AI-supported scholarly work.

4.2. Repositioning the Human: Editorial and Epistemic Imperatives

As large language models increasingly approximate scholarly prose, questions arise regarding their role in academic knowledge systems. The evidence synthesized in this review indicates that LLMs cannot function as autonomous epistemic agents because of their technical, ethical, and epistemic limitations. These systems lack contextual sensitivity, moral reasoning, and value alignment, and operate as probabilistic generators rather than accountable actors [9], [11], [16], [36]. Accordingly, human oversight is not a corrective add-on but rather a structural requirement. Within the Scholarly HI–AI Loop, human contributors act as curators, validators, and interpreters, filtering outputs, identifying errors or hallucinations, and aligning content with disciplinary and ethical standards [37]. This role is particularly

salient in fields shaped by interpretive and normative nuances, including the humanities, clinical ethics, and social sciences [38].

Empirical examples across the reviewed literature demonstrate that even advanced LLMs frequently require substantial human correction at various stages of the scholarly workflow. As summarized in Table 5, failures related to citation accuracy, authorship agency, and rhetorical coherence were reported from query expansion through the final proofreading. Across these cases, the maintenance of academic integrity depended on active human involvement. These observations reinforce the need to position humans as embedded epistemic actors, rather than passive validators of AI output.

Table 5
Use cases and pitfalls in AI-assisted academic writing across the HI–AI Loop

HI–AI Loop Stage	Human Oversight Required For	Observed AI Issue or Success	Reference
1. Query Expansion	Domain-specific search term refinement and contextual relevance filtering	LLMs enhanced recall but lacked sensitivity to discipline-specific jargon and emergent topics	[10]
2. Auto-Screening	Validation of inclusion/exclusion decisions and semantic clustering	AI retrieved high-relevance abstracts but failed to identify critical niche sources	[21]
3. Draft Summary	Identification of hallucinated statements, source misrepresentation, and authorship agency	AI-generated summaries included fabricated citations and, in some cases, displaced human authorship through passive adoption	[7]
4. Fact-Checking	Correction of scientific errors and restoration of factual integrity	Human reviewers substantially revised AI outputs to meet editorial and domain standards	[7]
5. Style Tuning	Restoration of scholarly tone, argumentative flow, and disciplinary nuance	LLMs produced grammatically smooth but rhetorically uniform and depersonalized prose	[26]
6. Citation Scan	Verification of citation presence, relevance, and accuracy	Tools like Elicit missed key references and occasionally introduced uncited material	[4]
7. Proof Polish	Final integration, ethical framing, and preservation of reviewer agency	MetaWriter accelerated review drafting but raised concerns about over-reliance and authorship	[39]

4.3. Reimagining Authorship and Attribution

The increasing capacity of GenAI to produce fluent academic prose complicates the existing definitions of authorship and plagiarism. Ethical uncertainty arises when AI-assisted paraphrasing occurs without disclosure or attribution, risking the misrepresentation of machine-generated content as human insight [14], [15], [40]. Questions of intellectual ownership are similarly unsettled, particularly for hybrid outputs co-produced by human users and models trained on large-scale corpora [26].

To address these challenges, codified attribution protocols are required that specify the role of AI (e.g., drafting, summarization, citation management), delineate the division of labor, and mandate structured disclosure, analogous to conflict-of-interest reporting. As shown in Table 2, current practices remain inconsistent across publishers and disciplines. Ethics-by-design mechanisms, such as interaction logs, citation audits, and authorship checklists, can further support accountability by making otherwise invisible contributions explicit. This reframing shifts the emphasis from output ownership to process transparency, recognizing AI as a contributory but non-moral agent in knowledge production.

4.4. Is the Traditional Literature Review Dead? Or just changing

The increasing use of LLMs in screening and summarizing the literature raises questions about the evolving role of traditional literature reviews. Fully automated approaches risk reducing synthesis to surface-level extraction, with limited attention to interpretive framing, theoretical depth or ethical reflexivity [40]. The evidence reviewed here indicates that even fluent AI-generated summaries frequently lack conceptual nuance and overlook minority or non-dominant perspectives [4], [10]. Hybrid models, such as the Scholarly HI–AI Loop, retain the interpretive strengths of human scholarship while leveraging AI for scalability. Rather than displacing the literature review, AI reshapes it into a reflective, context-aware synthesis. This perspective aligns with broader critiques highlighting the risk of epistemic injustice when generative systems privilege dominant paradigms and marginalize alternative epistemologies [38].

The reviewed evidence suggests that literature reviews remain fundamentally an interpretive scholarly practice. Operationalizing the HI–AI Loop translates its stages into concrete responsibilities for authors, editors, reviewers, and institutions, embedding accountability across the review life cycle. Such models position responsible AI adoption as a normative commitment, grounded in fairness, inclusivity, and institutional trust.

4.5. Designing Human–AI Collaboration for Trust and Value

Evidence from education, creative writing, and human–machine interaction suggests that the quality of human–AI collaboration depends on clear role delineation, iterative interaction, and a shared understanding of system limitations. Outcomes improve when humans focus on interpretation and contextualization, while AI supports speed, pattern recognition, and draft generation [19]. In contrast, one-shot prompting approaches are frequently associated with generic or weakly grounded outputs.

Therefore, trust in hybrid workflows is anchored in transparent processes rather than output fluency alone [41]. Collaborative literacy—encompassing prompt design, critical interpretation of AI responses, and reflective revision— has emerged as a necessary skill set for researchers [26], [38], [39], [42]. These findings point to a broader shift from viewing AI as a utility to recognizing it as a cognitive partner embedded within epistemic workflows [26], [35].

4.6. Stakeholder Guidance and Real-World Readiness

The Scholarly HI–AI Loop is intended as a practical framework for embedding responsible AI collaboration in the scholarly publishing ecosystem. Its implementation requires distributed accountability across researchers, editors, reviewers, institutions, publishers, ethical bodies, readers and tool developers. Researchers are primarily responsible for ensuring interpretive precision, validating AI-generated outputs, and providing accurate attribution [4], [26]. Editors and peer reviewers play crucial roles in enforcing disclosure standards and evaluating AI-generated content [16], [22], [33].

Institutional support is especially important in low- and middle-income countries (LMICs), where differences in infrastructure and training can worsen existing inequities [12], [27]. Therefore, publishers must take the lead in establishing standardized policies regarding AI disclosure and authorship guidelines [14], [15], [30]. Developers should prioritize explainability, bias reduction, and auditability in tool design [19]. Ethical oversight bodies, readers, and funders further contribute by defining the thresholds for acceptable use and supporting inclusive access. The stakeholder responsibilities aligned with the HI–AI Loop are summarized in Table 6.

Table 6
Governance recommendations for stakeholders in AI-assisted scholarly publishing

Stakeholder	Key Responsibilities	References
Researchers and Authors	Apply hybrid human–AI workflows with critical oversight; validate AI-generated content; ensure narrative coherence; disclose AI use; and maintain responsible attribution.	[4], [14], [26]
Editors	Enforce AI disclosure policies; incorporate governance checklists into editorial and peer-review processes; and clarify the respective roles of authors and AI systems to maintain scholarly integrity.	[16], [33]
Peer Reviewers	Assess the transparency and quality of AI-assisted content; identify hallucinations, bias, and unverifiable citations; and evaluate compliance with AI disclosure requirements.	[16], [22], [38]
Institutions	Provide training in responsible human–AI collaboration; establish institutional AI governance policies; and promote equitable access to trustworthy AI tools, particularly in low- and middle-income country settings.	[12], [27], [34]
Publishers and Journals	Standardize editorial policies on AI use; require transparent disclosure of AI assistance; and define acceptable boundaries for AI contributions in scholarly publishing.	[14], [15], [30]
Readers and Knowledge Users	Critically evaluate AI-assisted scholarly outputs; expect transparent disclosure of AI use; and report inaccuracies, citation errors, or concerns regarding research integrity.	[25], [28]
Ethics Oversight Bodies	Develop guidance for responsible AI use in scholarly communication; clarify authorship boundaries; establish audit responsibilities; and support ethical governance of AI-assisted research.	[18], [22]
AI Tool Developers and Platforms	Design transparent, explainable, and human-centered AI systems; mitigate algorithmic bias; and provide audit trails and governance features to support responsible scholarly use.	[19]
Funders and Policy Agencies	Require AI transparency and governance plans in funded research; support equitable access to trustworthy AI infrastructure; and promote responsible AI innovation through policy and funding mechanisms.	[12], [28]

Note: *The recommendations are aligned with the seven stages of the proposed Scholarly HI–AI Loop: scoping, retrieval, synthesis, validation, narrative construction, disclosure, and ethical governance.*

4.7. Co-Producing Scholarly Integrity in the AI Era

The integration of GenAI into academic practices demands a reevaluation of authorship, peer review, and accountability. The Scholarly HI–AI Loop provides a governance structure that emphasizes coproduction, shared cognition, and editorial responsibility. This approach, rather than replacing established academic principles, underscores the importance of human judgment in maintaining scholarly trust, thereby rejecting dichotomies such as automation and authenticity. This model reflects broader transformations within knowledge-intensive fields, where AI-driven methodologies are reshaping

scholarly workflows, decision-making, and standards of ethical responsibility [43]. The practical integration of ethical AI requires the establishment of comprehensive policies, training initiatives, and governance frameworks that involve all relevant stakeholders [17]. Consequently, ethical co-production is a fundamental requirement, not a discretionary element. [36]. The HI–AI Loop provides a structured approach for maintaining meaning-making responsibility with human scholars as AI accelerates the production of knowledge.

4.8. Limitations and Future Directions

This narrative review and the diversity of the included studies precluded a quantitative synthesis. The generalizability of the findings is constrained by variations in the AI tools evaluated, disciplinary contexts, and methodological approaches used in the included studies. Furthermore, restricting the review to English-language publications may have introduced cultural and linguistic biases. The rapid evolution of AI technologies and the continual refinement of editorial and institutional policies also mean that some practices described in this review may change. Importantly, the proposed Scholarly HI–AI Loop is a conceptually derived governance framework informed by synthesized literature rather than an empirically validated model. Consequently, future implementation and evaluation studies are needed to assess its feasibility, usability, effectiveness, and impact across different disciplines and institutional contexts.

Future research should also investigate hybrid human–AI workflows using standardized evaluation metrics for accuracy, transparency, interpretive quality, and ethical accountability. The development of discipline-specific ethical checklists and real-world implementation studies could further support consistent evaluation. Moreover, advancing ethical AI literacy requires training programs that integrate technical competencies with critical analytical thinking. As human–AI collaboration in academic writing becomes increasingly common, institutions must establish robust and enforceable policies governing disclosure, fairness, accountability, and peer-review standards to sustain trustworthiness and inclusivity in scholarly communication.

5. Conclusion

This study examined the current use of AI-assisted systems in academic writing and developed an evidence-informed governance framework for their responsible integration. The synthesis of 38 peer-reviewed studies indicates that, although AI can effectively support literature screening, summarization, drafting, and organization, scholarly quality and research integrity depend on sustained human oversight. Across the reviewed literature, human judgment remains essential for contextual interpretation, critical evaluation, ethical reasoning, and accountability. In response to these findings, this study proposes the Scholarly HI–AI Loop, a seven-stage governance framework that formalizes human oversight points throughout the academic writing process. Rather than positioning AI as an autonomous epistemic agent, the framework embeds AI within a human-led workflow in which intellectual responsibility, ethical judgment, and final accountability rest with the human author. By translating governance principles into a structured scholarly workflow, the framework provides a practical approach to promote transparency, responsible AI use, and authorship integrity. More broadly, the findings shift the focus from AI capability

alone to the governance structures required for trustworthy human–AI collaborations. Future research should empirically evaluate the implementation, usability, and effectiveness of the Scholarly HI–AI Loop across disciplinary and institutional contexts, thereby strengthening evidence-based governance for responsible AI-assisted academic communication.

Declarations

Author Contribution

T.A. conceptualized the study. T.A., S.P., and K.M. designed the review methodology. T.A. conducted the literature search. T.A. and K.M. performed screening and data extraction. S.K. adjudicated disputed inclusion decisions. T.A., K.M., and S.P. conducted the evidence synthesis and developed the governance framework. S.P. and K.M. supervised the study. All authors (T.A., L.N., S.P., O.J.L., S.K., K.M., S.S., and S.R.) contributed to writing, review, and editing, and all authors read and approved the final manuscript.

Data Availability

All data supporting the findings of this study are derived from publicly available, published literature cited in the manuscript and reference list. No primary datasets were generated or analyzed during the current study.

References

1. J. de la Torre-López, A. Ramírez, and J. R. Romero, “Artificial intelligence to automate the systematic review of scientific literature,” *Computing*, vol. 105, pp. 2171–2194, 2023, doi: 10.1007/s00607-023-01181-x.
2. P. Tomczyk, P. Brüggemann, and D. Vrontis, “AI meets academia: Transforming systematic literature reviews,” *EuroMed J. Bus.*, vol. 21, no. 1, pp. 345–369, 2026, doi: 10.1108/EMJB-03-2024-0055.
3. Y. Li, S. Datta, M. Rastegar-Mojarad, K. Lee, H. Paek, J. Glasgow, *et al.*, “Enhancing systematic literature reviews with generative artificial intelligence: Development, applications, and performance evaluation,” *Journal of the American Medical Informatics Association*, vol. 32, no. 4, pp. 616–625, 2025, doi: 10.1093/jamia/ocaf030.
4. M. Mostafapour, J. H. Fortier, K. Pacheco, H. Murray, and G. Garber, “Evaluating literature reviews conducted by humans versus ChatGPT: Comparative study,” *JMIR AI*, vol. 3, e56537, 2024, doi: 10.2196/56537.
5. N. Silva and D. Wickramaarachchi, “Enhancing systematic literature reviews: Evaluating the performance of LLM-based tools across key systematic literature review stages,” in *Proc. 5th Int. Conf. Adv. Res. Comput. (ICARC)*, 2025, pp. 1–6, doi: 10.1109/ICARC64760.2025.10963273.

6. T. Susnjak, P. Hwang, N. H. Reyes, A. L. C. Barczak, T. R. McIntosh, and S. Ranathunga, “Automating research synthesis with domain-specific large language model fine-tuning,” *ACM Trans. Knowl. Discov. Data*, vol. 19, no. 3, Art. 68, 2025, doi: 10.1145/3715964.
7. S. Ahn, “The transformative impact of large language models on medical writing and publishing: Current applications, challenges and future directions,” *Korean J. Physiol. Pharmacol.*, vol. 28, no. 5, pp. 393–401, 2024, doi: 10.4196/kjpp.2024.28.5.393.
8. E. O. Arhin, F. C. Ainoo, and K. Amponsah, “Integration of artificial intelligence (AI) into academic literature reviews: An overview,” *World J. Adv. Sci. Technol.*, vol. 7, no. 1, pp. 6–23, 2025, doi: 10.53346/wjast.2025.7.1.0010.
9. A. D. Lindström, L. Methnani, L. Krause, P. Ericson, Í. Martínez de Rituerto de Troya, D. Coelho Mollo, and R. Dobbe, "Helpful, harmless, honest? Sociotechnical limits of AI alignment and safety through reinforcement learning from human feedback," *Ethics Inf. Technol.*, vol. 27, Art. 28, 2025, doi: 10.1007/s10676-025-09837-2.
10. D. Tosi, “Comparing generative AI literature reviews versus human-led systematic literature reviews: A case study on big data research,” *IEEE Access*, vol. 13, pp. 56210–56219, 2025, doi: 10.1109/ACCESS.2025.3554504.
11. A. Holzinger, K. Zatloukal, and H. Müller, “Is human oversight to AI systems still possible?,” *New Biotechnol.*, vol. 85, pp. 59–62, 2025, doi: 10.1016/j.nbt.2024.12.003.
12. F. Bolaños, A. Salatino, F. Osborne, and E. Motta, “Artificial intelligence for literature reviews: Opportunities and challenges,” *Artif. Intell. Rev.*, vol. 57, p. 259, 2024, doi: 10.1007/s10462-024-10902-3.
13. S. Gregor, “Responsible artificial intelligence and journal publishing,” *J. Assoc. Inf. Syst.*, vol. 25, no. 1, pp. 48–60, 2024, doi: 10.17705/1jais.00863.
14. S. J. Kim, “Research ethics and issues regarding the use of ChatGPT-like artificial intelligence platforms by authors and reviewers: A narrative review,” *Sci. Editing*, vol. 11, no. 1, pp. 1–8, 2024, doi: 10.6087/kcse.343.
15. Z. Koçak, “Publication ethics in the era of artificial intelligence,” *J. Korean Med. Sci.*, vol. 39, Art. e249, 2024, doi: 10.3346/jkms.2024.39.e249.
16. H. Sharma and M. Ruikar, “Artificial intelligence at the pen’s edge: Exploring the ethical quagmires in using artificial intelligence models like ChatGPT for assisted writing in biomedical research,” *Perspect. Clin. Res.*, vol. 15, no. 3, pp. 108–115, 2024, doi: 10.4103/picr.picr_196_23.
17. A. Bozkurt, “GenAI et al.: Cocreation, authorship, ownership, academic ethics and integrity in a time of generative AI,” *Open Praxis*, vol. 16, no. 1, pp. 1–10, 2024, doi: 10.55982/openpraxis.16.1.654.
18. S. Järvelä, A. Nguyen, and A. Hadwin, “Human and artificial intelligence collaboration for socially shared regulation in learning,” *Brit. J. Educ. Technol.*, vol. 54, no. 4, pp. 743–759, 2023, doi: 10.1111/bjet.13325.
19. Y. Liu *et al*., “Consolidating human–AI collaboration research in organizations: A literature review,” *J. Comput. Signal Syst. Res.*, vol. 4, no. 1, pp. 55–73, 2025, doi: 10.71222/7dehvd30.

20. E. Mosqueira-Rey, E. Hernández-Pereira, D. Alonso-Ríos, *et al*., “Human-in-the-loop machine learning: A state of the art,” *Artif. Intell. Rev.*, vol. 56, pp. 3005–3054, 2023, doi: 10.1007/s10462-022-10246-w.
21. S. Spillias, P. Tuohy, M. Andreotta, R. Annand-Jones, F. Boschetti, C. Cvitanovic, J. Duggan, E. A. Fulton, D. B. Karcher, C. Paris, R. Shellock, and R. Trebilco, “Human–AI collaboration to identify literature for evidence synthesis,” *Cell Rep. Sustain.*, vol. 1, no. 7, p. 100132, 2024, doi: 10.1016/j.crsus.2024.100132.
22. G. E. Kaebnick, D. C. Magnus, A. Kao, M. Hosseini, D. Resnik, V. Dubljević, C. Rentmeester, B. Gordijn, and M. J. Cherry, "Editors' statement on the responsible use of generative AI technologies in scholarly journal publishing," Med. Health Care Philos., vol. 26, pp. 499–503, 2023, doi: 10.1007/s11019-023-10176-6.
23. M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, *et al.*, “The PRISMA 2020 statement: An updated guideline for reporting systematic reviews,” *BMJ*, vol. 372, n71, 2021, doi: 10.1136/bmj.n71.
24. A. C. Tricco, E. Lillie, W. Zarin, K. K. O’Brien, H. Colquhoun, D. Levac, *et al.*, “PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation,” *Ann. Intern. Med.*, vol. 169, no. 7, pp. 467–473, 2018, doi: 10.7326/M18-0850.
25. H. Else, “Abstracts written by ChatGPT fool scientists,” *Nature*, vol. 613, no. 7944, p. 423, 2023, doi: 10.1038/d41586-023-00056-7.
26. J. Hutson, “Human–AI collaboration in writing: A multidimensional framework for creative and intellectual authorship,” *Int. J. Changes Educ.*, vol. 5, no. 2, Art. 908, 2025, doi: 10.47852/bonviewIJCE52024908.
27. T. Anthuvan and K. Maheshwari, "AI-C2C (conscious to conscience): A governance framework for ethical AI integration," *AI Ethics*, vol. 5, pp. 4241–4253, 2025, doi: 10.1007/s43681-025-00736-2.
28. F. V. Giarmoleo, I. Ferrero, M. Rocchi, and M. M. Pellegrini, “What ethics can say on artificial intelligence: Insights from a systematic literature review,” *Bus. Soc. Rev.*, vol. 129, no. 2, pp. 258–292, 2024, doi: 10.1111/basr.12336
29. S. Kalra and S. C. Garcha, *Drugs for Medical Management of Obesity*: *A Machine-Generated Literature Overview*. Singapore: Springer, 2025, doi: 10.1007/978-981-96-1651-0.
30. H. Mondal, S. Mondal, and J. K. Behera, “Artificial intelligence in academic writing: Insights from journal publishers’ guidelines,” *Perspect. Clin. Res.*, vol. 16, no. 1, pp. 56–57, 2025, doi: 10.4103/picr.picr_67_24.
31. O. J. Lakhani, “Artificial intelligence in diabetes management and research,” *Chron. Diabetes Res. Pract.*, vol. 3, no. 1, pp. 5–7, 2024, doi: 10.4103/cdrp.cdrp_14_23.
32. S. Kalra, O. Lakhani, and N. Kapoor, "Asimov's laws and person-centered artificial intelligence," *J. Assoc. Physicians India*, vol. 73, no. 6, pp. 11–12, 2025, doi: 10.59556/japi.73.1002.
33. T. I. Leung, T. de Azevedo Cardoso, A. Mavragani, and G. Eysenbach, "Best practices for using AI tools as an author, peer reviewer, or editor," *J. Med. Internet Res.*, vol. 25, e51584, 2023, doi: 10.2196/51584. PMID: 37651164, PMCID: 10502596.

34. J. Morley, et al., “The ethics of AI in health care: A mapping review,” *Soc. Sci. Med.*, vol. 260, p. 113172, 2020, doi: 10.1016/j.socscimed.2020.113172.
35. M. Puerta-Beldarrain, O. Gómez-Carmona, R. Sánchez-Corcuera, D. Casado-Mansilla, D. López-de-Ipiña, and L. Chen, “A multifaceted vision of the human–AI collaboration: A comprehensive review,” *IEEE Access*, vol. 13, pp. 29375–29405, 2025, doi: 10.1109/ACCESS.2025.3536095.
36. S. Knight, A. Shibani, and N. Vincent, “Ethical AI governance: Mapping a research ecosystem,” *AI Ethics*, vol. 5, no. 2, pp. 841–862, 2025, doi: 10.1007/s43681-023-00416-z.
37. A. Thurzo, “How is AI transforming medical research, education and practice?,” *Bratisl. Med. J.*, vol. 126, no. 3, pp. 243–248, 2025, doi: 10.1007/s44411-025-00063-2
38. A. Nguyen, Y. Hong, B. Dang, and X. Huang, “Human–AI collaboration patterns in AI-assisted academic writing,” *Stud. High. Educ.*, vol. 49, no. 5, pp. 847–864, 2024, doi: 10.1080/03075079.2024.2323593.
39. L. Sun, S. Tao, J. Hu, and S. P. Dow, “MetaWriter: Exploring the potential and perils of AI writing support in scientific peer review,” *Proc. ACM Hum.-Comput. Interact.*, vol. 8, no. CSCW1, pp. 1–32, 2024, doi: 10.1145/3637371.
40. A. Skulmowski and P. Engel-Hermann, “The ethics of erroneous AI-generated scientific figures,” *Ethics Inf. Technol.*, vol. 27, Art. 31, 2025, doi: 10.1007/s10676-025-09835-4.
41. S. Mishra, "Artificial intelligence: A review of progress and prospects in medicine and healthcare," *J. Electron. Electromed. Eng. Med. Inform.*, vol. 4, no. 1, pp. 1–23, Jan. 2022, doi: 10.35882/jeeemi.v4i1.1.
42. Z. B. Akhtar and A. T. Rawol, “Biomedical engineering (BME) perspectives exploration for investigation analysis: Artificial intelligence (AI) and extended reality (XR),” *Indones. J. Electron. Electromed. Eng. Med. Inform.*, vol. 6, no. 3, pp. 132–146, Aug. 2024, doi: 10.35882/jeeemi.v6i3.4.
43. T. Anthuvan, S. Prabhuram, G. Raju, K. Maheshwari, and A. Mishra, “Bring Your Own AI (BYOAI) in the Workplace: Patterns, Pitfalls, and the BYOAI-Gov™ Framework for Behavior-Aware Governance,” *Lex Localis – Journal of Local Self-Government*, vol. 23, no. S6, pp. 2693–2716, 2025, doi: 10.52152/mmrgh422.

Figures

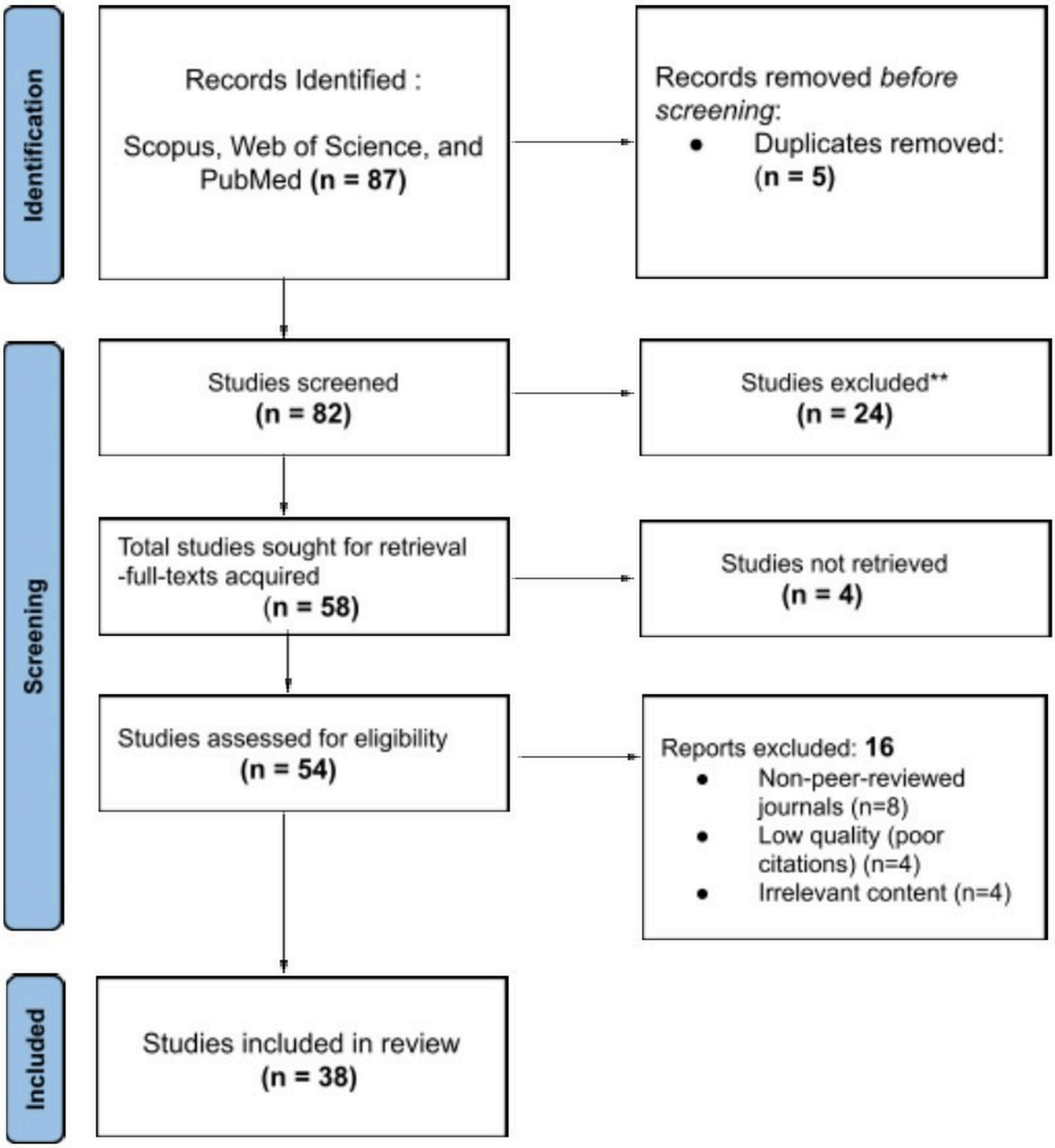


Figure 1

PRISMA 2020 study selection flow diagram [23], [24].

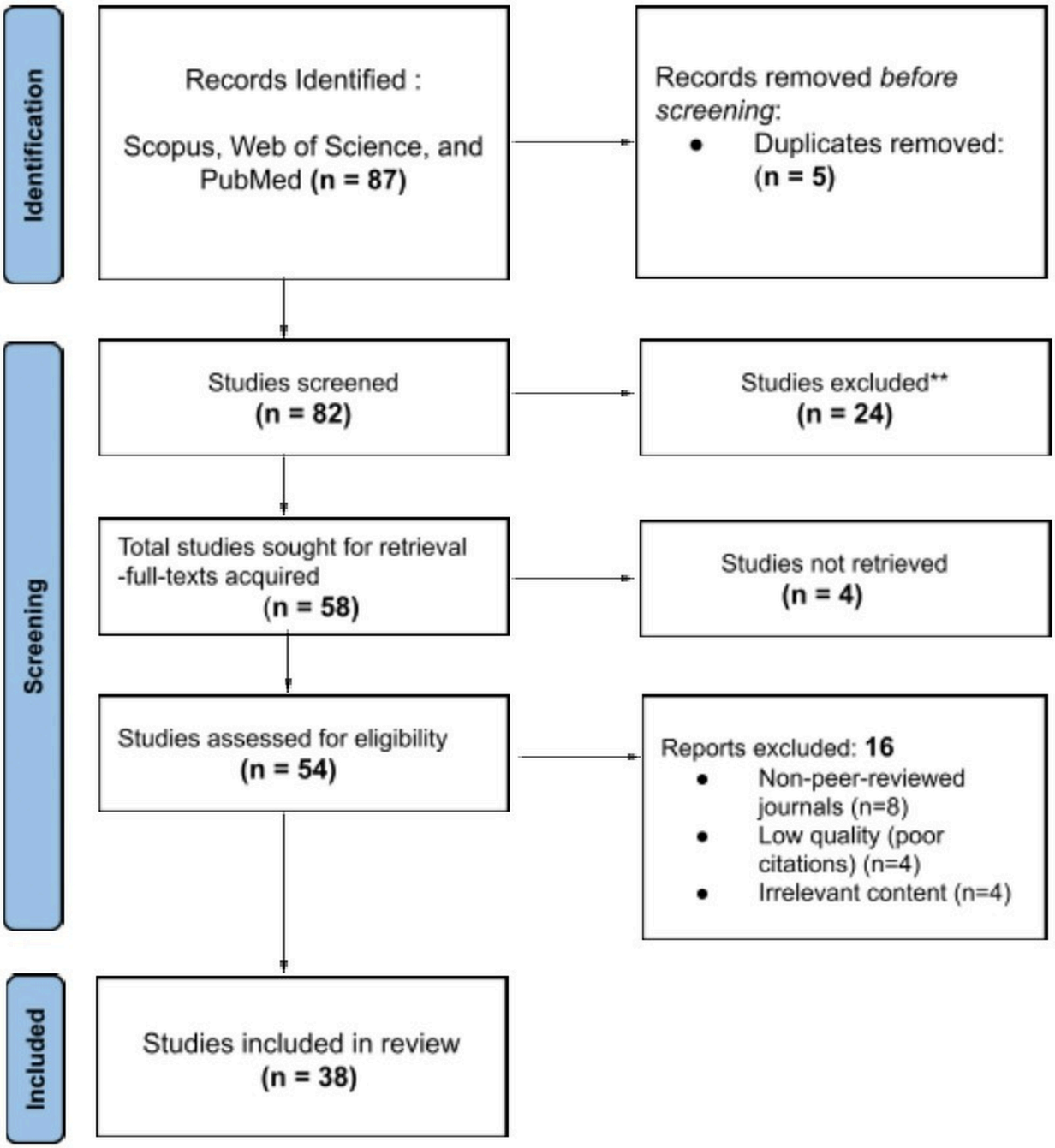


Figure 1

PRISMA 2020 study selection flow diagram [23], [24].

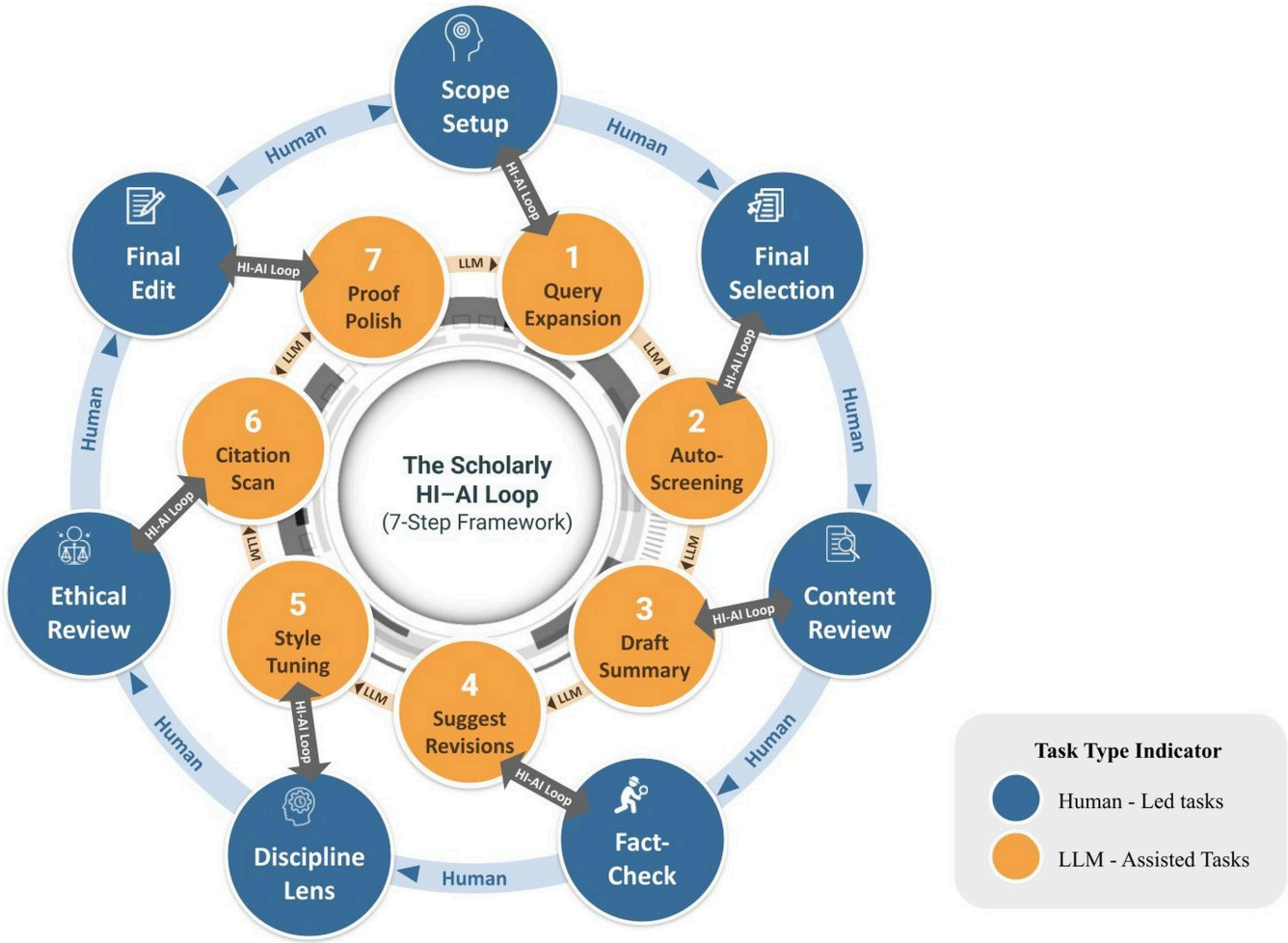


Figure 2

The proposed Scholarly HI–AI Loop framework for AI-assisted academic writing.

Note: A seven-stage framework for hybrid human–AI collaboration in academic writing and review workflows. LLM-assisted tasks (in orange) include query expansion, summarization, and revision suggestions; human-led stages (in blue) ensure contextual, disciplinary, and ethical integrity throughout the process.

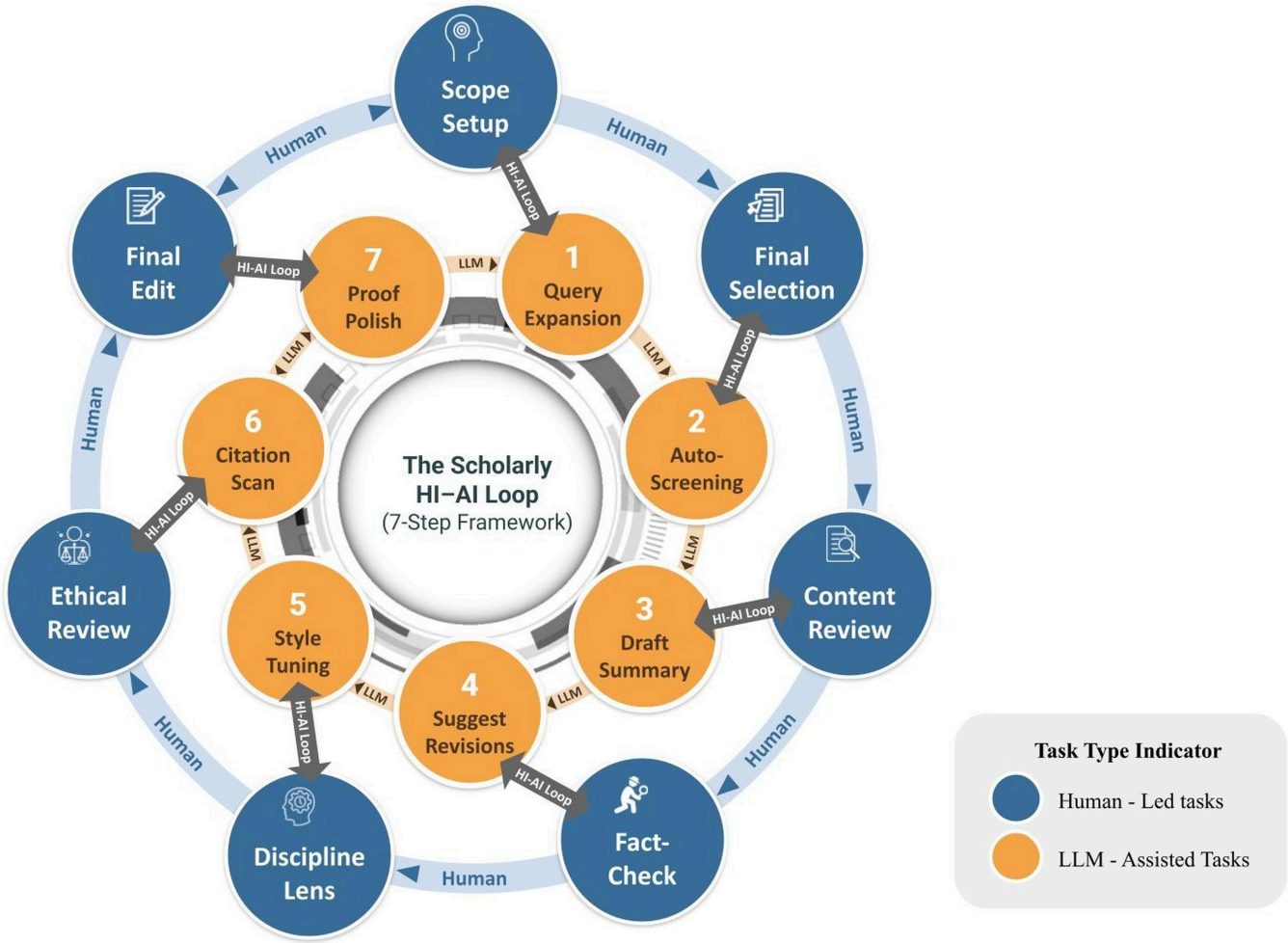


Figure 2

The proposed Scholarly HI–AI Loop framework for AI-assisted academic writing.

Note: A seven-stage framework for hybrid human–AI collaboration in academic writing and review workflows. LLM-assisted tasks (in orange) include query expansion, summarization, and revision suggestions; human-led stages (in blue) ensure contextual, disciplinary, and ethical integrity throughout the process.