

Viewpoint

# From AI Disclosure to Judgment Traceability: Addressing Research Waste in AI-Assisted Health Sciences Publishing

Tino Hsieh<sup>1,2</sup>, MSN; Ya-Ting Chen<sup>2</sup>, BSN

<sup>1</sup>School of Nursing, National Yang Ming Chiao Tung University, Taipei, Taiwan

<sup>2</sup>Department of Emergency Medicine, National Taiwan University Hospital, Taipei, Taiwan

**Corresponding Author:**

Tino Hsieh, MSN

School of Nursing

National Yang Ming Chiao Tung University

No 155, Section 2, Linong Street

Taipei, 112

Taiwan

Phone: 886 966 737 862

Email: [t22671884@gmail.com](mailto:t22671884@gmail.com)

## Abstract

Generative artificial intelligence (GAI) is changing how health sciences manuscripts are produced, revised, and evaluated. Publication guidance emphasizes necessary safeguards, such as disclosure, authorship, confidentiality, transparency, verification, and human responsibility, for AI-assisted content. However, disclosure alone does not establish whether consequential claims in a manuscript are grounded in accountable authorial and domain judgment. This viewpoint addresses this governance gap. We define expression-rich but judgment-thin scholarship as AI-amplified scholarly output in which linguistic fluency, structural coherence, and surface plausibility exceed the traceability of the reasoning, evidence interpretation, uncertainty, contextual limits, and responsibility behind major claims. We frame AI-induced research waste as a conceptual extension of the established research waste literature rather than as a measured prevalence claim. The concern is not that AI-assisted writing is inherently low quality; rather, GAI can make it easier to produce large volumes of polished text whose underlying judgments may be weakly supported, insufficiently limited, or difficult to defend. Building on emerging work on claim accountability, provenance, responsible AI-assisted writing, and publication ethics, we propose domain judgment traceability as a claim-centered governance approach for health sciences publishing. The framework asks whether major clinical, methodological, statistical, ethical, policy, and implementation claims can be traced to an appropriate evidence base, an identifiable interpretive step, explicit uncertainty, population and setting limits, equity and harm considerations, proportionate recommendation strength, and accountable human authorship. This approach shifts the unit of governance from the AI tool or prompt to the major scholarly claim. It also distinguishes disclosure from defensibility: disclosure describes whether and how AI contributed to manuscript development, whereas judgment traceability examines whether the resulting claim can be reconstructed, challenged, and defended by responsible authors. We propose practical implications for authors, reviewers, editors, and journals, including claim-level verification, proportionate requests for clarification, and selective piloting of traceability prompts rather than routine AI detection or exhaustive prompt surveillance. We also identify safeguards against gatekeeping, including attention to language equity, reviewer subjectivity, and the risk of conflating linguistic polish with scholarly competence. Domain judgment traceability is presented as a provisional conceptual governance aid, not a validated scoring instrument or reporting guideline. Its purpose is to complement—not replace—existing disclosure policies and reporting standards by making accountable judgment more visible where claims may influence clinical care, research interpretation, policy, implementation, or public trust. The framework creates an empirical agenda: future studies can test whether claim-centered prompts improve reviewer agreement, citation accuracy, proportionality between evidence and recommendations, author experience, and editorial usefulness without imposing an excessive documentation burden. The goal is not to reject AI-assisted scholarship or require disclosure of every prompt but to ensure that increasingly scalable scholarly expression is proportionate, defensible, context sensitive, and answerable to human judgment in health sciences publishing.

(*J Meta and Res Int* 2026;1:e106068) doi: [10.2196/106068](https://doi.org/10.2196/106068)

**KEYWORDS**

generative artificial intelligence; artificial intelligence; AI; health sciences publishing; research integrity; metascience; scientific publishing; research waste; domain judgment; publication ethics; AI disclosure; scholarly communication

**Background**

Generative AI (GAI) has rapidly entered scholarly writing, with large-scale lexical analysis suggesting detectable traces of large language model (LLM)–assisted writing in biomedical abstracts [1]. In health sciences publishing, AI-assisted tools may be used to organize manuscripts, improve clarity, support literature navigation, identify possible omissions, and reduce language-related barriers for authors who work in an English-dominant publication environment [2-4]. These uses are not inherently problematic; peer-reviewed commentary has emphasized both potential productivity and accessibility benefits of GAI and the need for transparent, verifiable human oversight [3-5]. The research integrity concern is therefore more specific than AI use itself: fluent and well-structured AI-assisted prose may make a manuscript appear complete without showing whether its major claims are accurate, proportionate to the evidence, contextually limited, and defensible by responsible authors [6-9].

Current publishing guidance has therefore focused on disclosure, authorship, confidentiality, transparency, verification, and human responsibility. Peer-reviewed guidance and policy analyses consistently state that AI systems cannot assume authorship responsibility and that human authors remain accountable for the accuracy, integrity, and transparency of the final work [3-5,9-14]. Policy analyses also show substantial heterogeneity across journals and publishers in allowable uses, disclosure expectations, and required descriptions of human oversight [3,13,14]. This heterogeneity matters because a disclosure statement may document the presence of AI without revealing whether an author has verified the evidence, reconstructed the reasoning behind a consequential claim, or calibrated the strength of a recommendation. The emerging literature therefore increasingly distinguishes procedural transparency from substantive accountability [4,7-9,15].

Building on emerging arguments that responsible AI governance should move beyond tool disclosure toward accountability for scholarly claims [7], this viewpoint asks what claim accountability requires in health sciences publishing. It defines expression-rich but judgment-thin scholarship, distinguishes it from adjacent research integrity problems, and proposes domain judgment traceability as a claim-centered governance approach for AI-assisted health sciences manuscripts. The intended audience includes journal editors and reviewers, health sciences authors, publication ethics committees, research integrity officers, and institutions developing AI use policies. Although the argument is especially relevant to clinically oriented manuscripts, it applies more broadly to public health, health informatics, health services research, implementation science, epidemiology, policy research, qualitative inquiry, and biomedical scholarship. Clinical judgment traceability is therefore treated as one important subtype within the broader concept of domain judgment traceability. This formulation is consistent with recent peer-reviewed arguments that responsible

governance should preserve provenance, verification, and identifiable human responsibility rather than treating disclosure or prohibition as sufficient ends in themselves [4,7-9].

This viewpoint does not claim that AI-assisted writing is inherently defective or that disclosure policies are unimportant. Instead, it identifies a narrower governance gap. Disclosure documents whether and how a tool contributed to manuscript preparation; judgment traceability asks whether the resulting claims are sufficiently supported, limited, and accountable. This distinction allows the manuscript to engage AI-related risks while preserving a balanced position toward legitimate AI-assisted language, organization, and revision support [3-5,10]. Its primary contribution is to operationalize emerging claim accountability arguments for health sciences publishing through domain judgment traceability: a way of asking whether major scholarly claims remain connected to evidence, interpretation, uncertainty, context, and accountable authorship [7,8]. The framing deliberately avoids the stronger assertion that judgment traceability has already been empirically shown to reduce research waste. Instead, it treats judgment traceability as a provisional governance aid whose usefulness, burden, fairness, and effects on editorial decision-making require empirical evaluation [4,7,8].

**Ethics Approval and Informed Consent**

Ethics approval and informed consent were not applicable because this viewpoint did not involve human participants, personal data, or identifiable clinical information.

**Why Disclosure Is Necessary but Insufficient**

Disclosure is a necessary part of responsible AI-assisted publishing because it informs editors, reviewers, and readers that a nonhuman system contributed to manuscript development. Peer-reviewed guidance increasingly recommends disclosures that specify not only the tool but also the task, extent of assistance, verification process, and location of AI-supported work [3,10-14]. Frameworks such as GAIDeT (Generative Artificial Intelligence Delegation Taxonomy) further provide structured language for documenting which research or writing tasks were delegated to GAI while preserving human oversight [16]. These approaches strengthen transparency because “AI was used” can otherwise conceal materially different practices, ranging from grammar correction to literature summarization, conceptual brainstorming, data analysis, or recommendation drafting. However, disclosure is a description of process rather than a direct test of claim quality. A manuscript may transparently disclose AI use and still contain overextended interpretations, unsupported clinical implications, fabricated or erroneous citations, or recommendations whose strength exceeds the underlying evidence [17-19]. Conversely, AI may be used for language editing, organization, or perspective checking,

while authors preserve rigorous source interpretation, contextual limits, and responsibility for every substantive claim [3-5].

This focus on claim quality addresses a practical weakness of tool-centered governance. A disclosure statement can be accurate yet uninformative about whether a recommendation is justified by the cited evidence. For example, an author may disclose AI-assisted drafting or summarization but still fail to show why a limited observational finding warrants a broad clinical implication; conversely, an author may use AI to improve readability while maintaining rigorous source interpretation and cautious recommendations [4,7-9,15]. A citation should therefore not merely appear relevant: the cited source should directly support the claim attributed to it, and authors should be able to explain how the evidence justifies the level of inference or recommendation being made. This distinction is particularly important because disclosure practices remain heterogeneous across journals, disciplines, and publishers [3,13,14]. Some policies request only the name of a tool, whereas others ask authors to describe the task, location of use, or degree of human verification [3,10,13,14]. Disclosure can support transparency, but it remains procedural unless connected to the claims the manuscript asks readers to accept. A claim-centered standard therefore complements rather than competes with disclosure by asking whether consequential interpretations and recommendations remain attributable to responsible human judgment [7,8].

## *Expression-Rich but Judgment-Thin Scholarship*

Expression-rich but judgment-thin scholarship is defined as scholarly work in which the language, structure, and apparent completeness of a manuscript exceed the traceability of the judgment supporting its major claims. The concept does not refer simply to poor writing, weak scholarship, or the use of GAI. Rather, it identifies a mismatch between scholarly expression and accountable reasoning: a manuscript may read as fluent, balanced, clinically relevant, and well organized while leaving unclear how its central claims were framed, interpreted, limited, and made defensible by responsible authors. This concern is compatible with broader warnings that GAI can create persuasive or seemingly authoritative outputs without reliable epistemic grounding and that apparent fluency should not be mistaken for understanding [4,6,20]. It also aligns with recent claim accountability and provenance arguments that the central governance question is whether a human can reconstruct and defend the reasoning, evidence, and verification behind what enters the scholarly record [7,8].

Judgment thinness is best understood as a property of the relationship among text, authorial process, and claims. A manuscript can be judgment thin without AI because conventional academic writing can also be formulaic, overgeneralized, densely cited but poorly interpreted, or disconnected from clinical and methodological realities. A manuscript can also be AI assisted and judgment rich if authors use AI to improve expression, identify missing perspectives, challenge assumptions, or organize reasoning while retaining responsibility for the final claims [3-5,9]. AI may amplify

judgment thinness, not because it uniquely causes weak reasoning, but because it can increase the speed and scale at which plausible scholarly text is produced and can make weakly examined reasoning appear polished enough to pass superficial review. The risk is therefore not the presence of machine-generated language itself; it is the possibility that fluency masks unresolved evidentiary, methodological, or contextual problems that responsible authors have not adequately examined [6-8,19].

This concept overlaps with, but is not reducible to, several existing concerns. Hallucination concerns fabricated or inaccurate content, paper mill production concerns fraudulent or industrialized paper generation, automation bias concerns overreliance on machine output [21,22], and epistemic opacity concerns difficulty understanding how algorithmic systems produce outputs [23]. Paper mills represent a documented threat to research integrity [24,25], and work on AI in science has warned that AI systems may produce illusions of understanding [6]. Concerns about fluent but ungrounded language model output, including the “stochastic parrots” critique, are relevant background but do not by themselves define the claim-level failure mode addressed here [20]. Judgment thinness specifically concerns whether the reasoning behind major claims is visible, proportionate, and defensible.

We propose several observable warning signs that may prompt closer review: broad clinical or policy implications that exceed the evidence cited; generic recommendations that do not specify population, setting, resource, or equity limits; literature summaries that list findings without interpreting their strength or relevance; conclusions that convert limited evidence into action-oriented recommendations without explaining inferential steps; citations that exist but do not support the sentence in which they are placed; and statements that appear authoritative but cannot be defended by a responsible author when challenged [7,8,17-19]. These features should prompt clarification of claim-to-evidence reasoning, not accusations of AI use. They are intentionally framed as review prompts rather than diagnostic criteria because AI detection tools have documented limitations and may produce inequitable misclassification, including for nonnative English writers [2,26]. Nor should the prompts penalize concise writing. Their purpose is to identify where a manuscript may require clearer explanation of evidence strength, inferential steps, contextual limits, or authorial responsibility. In this sense, judgment traceability functions as a quality and accountability standard rather than an AI detection strategy [7,8].

## *From Research Waste to AI-Induced Research Waste*

Research waste is commonly discussed in relation to the production, conduct, reporting, dissemination, and usability of empirical research [27-29]. In health sciences, this framing has ethical as well as economic significance because low-value research can consume participant contributions, public funding, institutional resources, reviewer labor, editorial attention, and reader trust. Contemporary research integrity literature also documents how fraudulent publication systems and paper mills

can introduce large volumes of unreliable work into the scholarly record and create downstream burdens for journals and evidence synthesis [24,25,30]. This literature provides the conceptual foundation for asking whether AI-assisted scholarly production may create an additional pathway to low-value output while also requiring caution: AI-induced research waste is proposed here as a conceptual extension, not as a measured prevalence claim.

Empirical evidence supports the plausibility of this concern without proving its prevalence. Large-scale lexical analysis suggests detectable traces of LLM-assisted writing in biomedical abstracts [1]. Separate studies demonstrate risks, including fraudulent-looking AI-generated medical manuscripts, fabricated or erroneous AI-generated citations, and citation errors that can appear bibliographically plausible [17-19]. These findings do not show that AI-assisted manuscripts are inherently lower quality or judgment thin. They do show why authors, reviewers, and editors cannot rely on surface fluency, citation density, or procedural disclosure alone when evaluating whether a manuscript's claims add proportionate knowledge value. The problem is compounded when unreliable material enters later scholarship: empirical work on paper mill contamination has shown that retracted fraudulent papers can be incorporated into systematic reviews and continue to be cited even after retraction [30]. Although paper mills and AI-assisted writing are distinct phenomena, both illustrate why scholarly systems need safeguards that evaluate the integrity of the claim-source relationship rather than only the appearance of a completed manuscript [24,25,30].

On this basis, AI-induced research waste should be understood as a possible pathway by which AI-assisted scholarly outputs consume attention without adding proportionate epistemic, clinical, methodological, or social value. The concern includes obvious problems, such as fabricated references and unreliable summaries, but it also includes a subtler failure: manuscripts that are fluent and rhetorically complete yet leave reviewers and readers to reconstruct missing judgment behind their major claims [6-8]. This formulation is intended to prevent overclaiming. The available empirical evidence supports component risks (detectable AI-assisted writing, fabricated or erroneous citations, fraudulent-looking text, detector bias, automation-related overreliance, and broader threats to provenance), but it does not measure the prevalence of judgment-thin scholarship as an empirical category [1,2,6,8,17-19,21,22,26,31]. The proposed contribution is therefore conceptual and governance-oriented: it names a failure mode that may become more scalable under AI-assisted writing conditions and specifies what would need to be visible for major claims to remain accountable. This emphasis also avoids equating productivity with value; the ethical concern is not that AI makes writing faster but that increased output can become wasteful when the additional text does not carry commensurate evidentiary or interpretive value [9,27-29].

## *From Clinical Judgment to Domain Judgment Traceability*

Because health sciences publishing includes diverse forms of inquiry, clinical judgment traceability should not be treated as the only relevant standard. Evidence-based practice has long emphasized that clinical decisions require integration of the best-available evidence, professional expertise, and patient values [32]. Yet, manuscripts in public health, epidemiology, implementation science, health informatics, qualitative inquiry, health services research, policy research, and biomedical science may not make direct clinical recommendations while still depending on domain-specific judgment. Recent discussions of claim accountability and provenance similarly emphasize that the defensibility of scholarly claims depends on identifiable human responsibility for evidence selection, interpretation, and verification regardless of discipline [7,8]. For this reason, this viewpoint uses domain judgment traceability as the broader governance concept.

Domain judgment traceability refers to the ability of readers, reviewers, and editors to identify how responsible authors interpreted evidence, handled uncertainty, considered context, and justified the level of claim being made. Clinical judgment traceability is one subtype that becomes especially important when a manuscript makes claims about patient care, professional practice, clinical workflow, health system operations, or policy decisions with foreseeable implications for care. Other forms include methodological, statistical, ethical, policy, and implementation judgment. Judgment types describe the substantive domains in which authors exercise judgment, whereas the traceability domains in Table 1 describe what must be made visible for a major claim to remain accountable. Where prior claim accountability work establishes that a named human must be able to reconstruct and defend each scholarly claim [7], domain judgment traceability specifies what should be visible for that defense in health sciences manuscripts: the evidence base, author interpretation, uncertainty, population and setting context, equity and harms, recommendation strength, and accountable authorship. Its additional contribution is to connect those elements to health sciences-specific concerns about research waste; transferability; patient or population consequences; and the proportionality of clinical, policy, and implementation recommendations. This approach extends rather than replaces existing publication ethics expectations and reporting standards [4,5,7-14]. The CONSORT-AI (Consolidated Standards of Reporting Trials–Artificial Intelligence) extension and DECIDE-AI (Developmental and Exploratory Clinical Investigation of Decision Support Systems Driven by Artificial Intelligence), for example, emphasize intended use, clinical setting, human-AI interaction, input and output handling, implementation context, and safety evaluation [33,34]. Judgment traceability is not proposed as a replacement reporting guideline; rather, these reporting concerns illustrate why claims about AI-supported care, workflow, risk prediction, or implementation should make the evidence base, setting, uncertainty, harms, and level of action justified visible.

**Table 1.** Proposed claim-centered framework for domain judgment traceability in AI-assisted health sciences manuscripts.

| Traceability domain                      | What authors should make visible                                                                                                                         | Potential failure mode                                                                | Example of adequate traceability                                                                                                                                                   |
|------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Evidence base                            | Studies, guidance, data, or conceptual sources supporting the claim                                                                                      | Sources are cited but not interpreted; weak evidence supports strong claims.          | The recommendation is based on two observational studies and should be interpreted cautiously.                                                                                     |
| Author interpretation                    | How responsible authors moved from evidence to meaning and why the claim follows                                                                         | Findings are listed or paraphrased without an identifiable inferential step.          | We interpret these findings as relevant to workflow redesign because the observed effect concerns response timing while recognizing that alternative explanations remain possible. |
| Uncertainty                              | What remains uncertain, including limitations, competing explanations, and confidence in the inference                                                   | The manuscript presents a qualified finding as settled or omits material uncertainty. | The direction of effect is consistent, but uncertainty remains because evidence is heterogeneous and external validation is limited.                                               |
| Population, setting, and context         | Who the claim applies to, where it applies, and under what resource or workflow conditions                                                               | A narrow finding is generalized to all patients, settings, or systems.                | These implications apply primarily to adult EDs <sup>a</sup> with integrated electronic health record alert systems and defined response protocols.                                |
| Equity and harms                         | Potential differential effects, exclusions, burdens, or unintended consequences                                                                          | The manuscript assumes equal benefit across groups or ignores possible harms.         | Performance should be assessed across relevant demographic, clinical, language, and access-related subgroups, when appropriate.                                                    |
| Recommendation strength                  | Whether the recommended action is proportionate to the strength, consistency, and applicability of the evidence                                          | A tentative implication is written as a universal implementation mandate.             | Current evidence supports further evaluation in appropriately resourced settings, not immediate universal adoption.                                                                |
| Accountable authorship and AI assistance | Which authors are responsible for interpretation and recommendation; whether AI materially supported major claims; and how relevant outputs were checked | Responsibility is diffused into passive wording or AI-generated generality.           | AI assisted drafting of this section; the authors checked the cited sources, revised the inference, and accept responsibility for the final recommendation.                        |

<sup>a</sup>ED: emergency department.

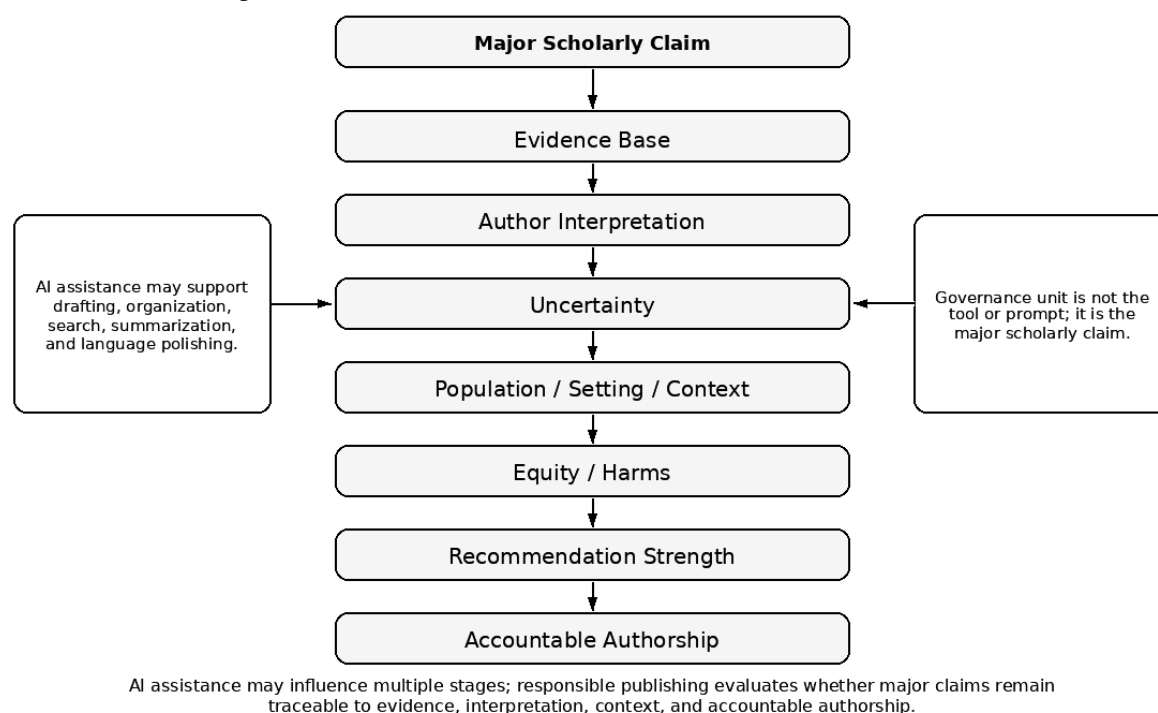
Different forms of domain judgment require different forms of traceability. Clinical judgment concerns whether claims about care, risk, diagnosis, treatment, triage, monitoring, or professional action are appropriate for the relevant population and setting. Methodological judgment concerns whether the study design, sampling, measurement, comparison, and interpretation are aligned with the question being asked. Statistical judgment concerns whether analytic choices, uncertainty, effect estimates, assumptions, and limitations are interpreted proportionately. Ethical judgment concerns whether benefits, burdens, risks, consent, privacy, equity, and potential harms are adequately considered. For the purposes of the framework proposed here, policy and implementation claims should additionally make feasibility, stakeholders, resources, workflows, workforce capacity, and unintended consequences visible rather than presenting these considerations as if they were automatically established by a generic AI reporting standard. The framework is therefore not restricted to bedside clinical decisions: A statistical claim may require traceability of analytic assumptions and uncertainty; an implementation claim may require traceability of staffing, infrastructure, and local feasibility; and a policy claim may require traceability of stakeholders, unintended consequences, and the level of action justified. These distinctions matter because health sciences

manuscripts often combine empirical evidence, interpretation, and recommendations within the same paper, and the strength of each layer should remain visible rather than being compressed into a single authoritative sentence [4,7,8].

### What Should Be Traceable?

Judgment traceability should be claim centered. It should apply primarily to major interpretive, clinical, methodological, ethical, policy, and implementation claims rather than to every sentence. A major claim is one that asks readers to accept an interpretation, change a practice, adopt a policy, cite a conclusion, prioritize an intervention, or treat a proposed implication, as warranted. Stronger claims require stronger traceability. Figure 1 therefore places the major scholarly claim, rather than the AI tool or prompt, at the center of governance. AI assistance may influence multiple stages of scholarly work, but the unit of governance is the claim and whether it remains traceable to an evidence base, author interpretation, uncertainty, population and setting context, equity and harms, recommendation strength, and accountable authorship [7,8]. This claim-centered orientation is also consistent with peer-reviewed calls to prioritize provenance and verification over attempts to govern GAI through prohibition alone [8].

**Figure 1.** Claim-centered judgment traceability in AI-assisted health sciences publishing. The figure presents a conceptual framework, not an empirically validated model. AI: artificial intelligence.



The practical value of a claim-centered approach is proportionality. A minor wording edit does not require extensive documentation, whereas a claim that implies a change in clinical practice, health policy, professional responsibility, or implementation priority should make its evidentiary basis and inferential pathway visible [4,7,8]. Before applying the traceability domains, authors and reviewers should first classify the claim as an evidence summary, inference, implication, recommendation, or speculation because the required level of support depends on what the statement asks readers to accept. Table 1 then translates judgment traceability into seven domains that correspond to Figure 1: evidence base, author interpretation, uncertainty, population and setting context, equity and harms, recommendation strength, and accountable authorship. These domains are not intended to create a scoring instrument; they provide a vocabulary for asking what must be visible for a major claim to be defensible. Existing reporting guidelines, such as CONSORT-AI and DECIDE-AI, demonstrate the broader value of making the intended use, setting, human-AI interaction, safety, and implementation context explicit, even though judgment traceability is not proposed as a replacement reporting guideline [33,34].

The framework also clarifies the difference between empirical support and conceptual contribution. Empirical studies can support component concerns, such as fabricated citations, detector bias, language barriers, publication policy heterogeneity, paper mill contamination, and automation bias [1-3,13,14,17-19,21,22,24-26,30,31]. They cannot, by themselves, prove the full construct of expression-rich but judgment-thin scholarship, because that construct is proposed here as an analytic lens. Accordingly, this viewpoint uses empirical evidence to support the plausibility and relevance of the problem while using explicit proposal language when advancing judgment traceability as a governance response [7,8].

This separation is important for metascience: a useful conceptual framework should generate testable questions without presenting its proposed constructs as though their prevalence, reliability, or effectiveness had already been established.

Table 1 summarizes the seven traceability domains shown in Figure 1. It is not a validated assessment instrument. Rather, it is a conceptual governance aid intended to help authors, reviewers, and editors identify what should be made visible for consequential scholarly claims to remain defensible.

### When AI Contributes an Idea: a Worked Example

A further question is whether a recommendation initially suggested by an AI tool can count as being supported by human judgment if the authors had not independently conceived it before the AI output was produced. The answer depends not on the origin of the idea but on the authorial work that follows. Accountability rests on whether responsible authors can reconstruct the reasoning, evaluate the evidence, consider alternative interpretations, identify limits, and defend the final claim as their own scholarly judgment [7-9]. This position is consistent with publication ethics arguments that AI systems cannot bear authorship responsibility and with peer-reviewed guidance that verification and human responsibility remain necessary even when GAI materially assists writing or idea development [4,5,9-12,14]. The practical distinction is between assistance that stimulates accountable scholarly work and delegation that leaves no responsible human able to defend the resulting claim.

This issue is also relevant to honest AI disclosure. If authors believe that acknowledging AI use will automatically trigger disproportionate procedural burdens, they may be less willing

to disclose assistance; conversely, generic disclosures may become performative if they are treated as sufficient evidence of responsible use [3,4,7,10]. A claim-centered approach reduces that risk by focusing scrutiny on the consequences of the claim rather than the mere presence of AI support. The same standard can therefore be applied to AI-assisted and non-AI-assisted manuscripts: are the major claims proportionate, source grounded, context sensitive, and attributable to accountable authors? This preserves the transparency value of disclosure while avoiding a governance model in which every prompt or stylistic edit receives the same scrutiny as a recommendation that could influence care, policy, or research priorities [7,8].

The worked example here illustrates this distinction in a clinical context, but the same logic can be adapted to other health sciences claims. We propose that a qualitative manuscript should make visible how interpretations were derived from data and context; an informatics manuscript should clarify intended users, workflow assumptions, model limitations, and potential harms; and a policy viewpoint should distinguish evidence supporting a proposed action from normative or speculative elements. These are applications of the proposed traceability framework rather than claims that one existing reporting guideline already covers all paper types [7,8]. Consider a hypothetical AI-assisted manuscript discussing an emergency department (ED) sepsis alert intervention. A judgment-thin implication might state, “AI-enabled sepsis alerts should be implemented across EDs to improve early recognition, accelerate treatment, and enhance patient safety.” This sentence is fluent and clinically plausible, but it does not show whether the supporting evidence came from adult or pediatric populations, academic or community hospitals, high-resource or underresourced settings, or prospective trials or early-stage evaluations or whether the alert system’s positive predictive value, false-positive burden, workflow effects, and performance across patient subgroups were adequately assessed. Empirical sepsis alert literature illustrates why these details matter: ED alert systems vary substantially across studies, and recent evidence shows both potential clinical benefits and persistent concerns about the alert burden, generalizability, and sociodemographic performance [35-37].

A judgment-traceable version would make the claim more proportionate: “Available evidence may support further evaluation of AI-enabled sepsis alerts in EDs with sufficient informatics infrastructure; clinician response protocols; and monitoring for false positives, alert burdens, workflow effects, and subgroup performance. The current evidence should not be interpreted as a universal implementation mandate. Before adoption, journals, institutions, and clinical leaders should examine whether the evaluated population, setting, workflow, AI system version, input data, output interpretation, and outcome measures correspond to the intended local use case.” The difference is not merely tone. The judgment-traceable version distinguishes evidence from inference and recommendation, identifies contextual limits, makes uncertainty visible, and preserves accountability for the level of action being proposed [7,8,33-37]. It also demonstrates why judgment traceability is not reducible to adding more citations: the same evidence base

can support very different claims, depending on how authors interpret scope, uncertainty, harms, and transferability.

## *Implications for Authors, Reviewers, Editors, and Journals*

For authors, judgment traceability means identifying major claims and ensuring that each is supported by an appropriate evidence base, an explicit interpretive step, and a proportionate level of recommendation. Authors should distinguish evidence summaries from inferences, implications, recommendations, and speculation and should verify that cited sources actually support the statements attributed to them [7,8,18,19]. Transparent AI-assisted writing can support this process when authors document human intellectual contribution and verification without turning governance into indiscriminate prompt-level surveillance [3,4,9,10,14-16]. The goal is not to record every prompt but to ensure that claims likely to influence interpretation, practice, policy, or implementation can be reconstructed and defended by responsible authors. A useful author-facing question is therefore not “Did AI write this sentence?” but “If challenged, can we show why this claim follows from the cited evidence, why its limits are appropriate, and who accepts responsibility for it?” [7,8].

For reviewers and editors, judgment traceability should function as an interpretive aid rather than a punitive label. Reviewers should not be asked to determine whether text is AI generated, particularly because detection tools may be unreliable and may disadvantage nonnative English writers [2,26]. Instead, they can examine whether cited evidence directly supports major claims, whether population and setting limits are specified, whether competing interpretations are acknowledged, and whether recommendation strength is proportionate to the evidence [7,8]. Editors can use this approach selectively when claims appear clinically, methodologically, ethically, or policy relevant but insufficiently supported. Journals could pilot structured traceability prompts before treating them as settled standards and evaluate outcomes, such as author burden, reviewer agreement, editorial usefulness, citation accuracy, correction requests, and perceived fairness. The staged approach also clarifies what journals should not routinely require: complete prompt histories, private drafting records, or speculative reconstructions of every AI interaction may be burdensome, difficult to standardize, and poorly aligned with the integrity risk [3,4,8,10]. A more feasible approach is to request clarification when a consequential claim lacks adequate evidence, context, uncertainty, or accountability, preserving the integrity function of peer review while avoiding a governance model that rewards paperwork over reasoning.

For journals and publishers, a practical starting point is not to require a new universal form but to identify where current editorial workflows already ask for evidence, limitation, and accountability. Traceability prompts could be integrated selectively into reviewer forms, revision letters, or editorial checks for manuscripts containing high-consequence clinical, policy, or implementation claims. For example, reviewers could be asked whether a major recommendation is directly supported by the cited evidence, whether the target population and setting

are explicit, and whether uncertainty or foreseeable harms have been acknowledged. Such prompts would operationalize claim accountability without requiring editors to infer how much text was generated by AI or to police every stage of drafting [7,8]. Existing disclosure taxonomies and peer-reviewed guidance can continue to document the role of AI in manuscript production, while judgment traceability addresses the separate question of whether the resulting claims are defensible [3,4,10,14,16]. Importantly, implementation should be evaluated rather than assumed beneficial. Journals would need to monitor whether additional prompts improve citation accuracy, reviewer consistency, and recommendation proportionality and whether they create new burdens, delays, or inequities for authors and reviewers. A governance intervention that increases paperwork without improving claim quality would itself risk becoming a form of editorial waste. The proposed framework therefore favors targeted clarification of consequential claims over blanket expansion of submission requirements, and it treats feasibility, fairness, and workload as empirical outcomes for future testing.

### *Equity, Subjectivity, and Gatekeeping*

Any framework that asks reviewers to evaluate judgment carries risks of subjectivity. Reviewers may mistake familiar disciplinary language, institutional prestige, seniority, native English fluency, or agreement with their own clinical preferences for stronger judgment. Evidence from peer-reviewed research suggests that author and affiliation information can influence reviewer behavior in some settings [31]. Research on AI text detection raises a separate equity concern: GPT detectors have misclassified nonnative English writing as AI generated in some evaluation settings [26], while nonnative English-speaking researchers face measurable burdens in reading, writing, publishing, and presenting science [2]. Domain judgment traceability should therefore be applied through observable claim features rather than impressions of author competence. A judgment-rich manuscript does not need ornate prose or extensive rhetorical display; it needs alignment among claim, evidence, inference, uncertainty, context, and accountability [7,8]. Authors should have opportunities to clarify allegedly insufficiently traceable claims, and reviewer concerns should be framed as requests for claim evidence clarification rather than accusations about AI use or author capability. These safeguards are essential if claim-centered governance is to improve accountability without reproducing linguistic or institutional gatekeeping [2,3,26,31].

### *Boundaries of Judgment Traceability and Future Research Priorities*

Judgment traceability should be understood as a proposed conceptual governance aid rather than a validated assessment instrument. Its domains and examples require refinement through stakeholder consultation, piloting, and empirical evaluation. Future work should examine whether claim-centered traceability improves reviewer agreement, author experience, editorial usefulness, citation accuracy, correction rates, and proportionality between evidence and recommendations while also measuring burden and unintended consequences [7,8].

Validation should proceed in stages: Qualitative work can first examine how authors, reviewers, and editors understand traceability; consensus methods can then refine candidate domains; pilot studies can assess interviewer agreement and feasibility; and comparative evaluations can test whether traceability prompts improve claim-source alignment or recommendation proportionality without creating inequitable barriers. The emerging literature on claim accountability, provenance, and responsible AI-assisted writing is developing rapidly, so the contribution of judgment traceability should be interpreted within this evolving conversation rather than as a claim of conceptual exclusivity [4,7,8,15].

The concept should also be applied proportionately across paper types. A clinical guideline, empirical study, qualitative analysis, policy commentary, laboratory report, implementation report, and philosophical essay make different kinds of claims and require different levels of traceability. Judgment traceability should not become a universal checklist applied mechanically to all scholarly writing, nor should it be used as a proxy for seniority, institutional prestige, English fluency, or disciplinary conformity [2,26,31]. Its purpose is to make consequential scholarly claims more accountable while preserving equitable access to legitimate AI-assisted language and writing support [2-4]. A boundary-focused implementation should also distinguish private reasoning from accountable scholarly justification. Judgment traceability does not require authors to reveal private chain-of-thought, informal notes, or every intermediate draft; it requires that consequential scholarly judgments—such as why evidence was interpreted in a particular way, why a recommendation is limited to a certain population or setting, and who accepts responsibility for the final claim—be sufficiently visible for review, revision, and accountability [7-9]. This boundary is important because responsible publishing needs verifiable justification, not surveillance of cognition.

### *Conclusion*

GAI does not make scholarship irresponsible by default; it may improve clarity, accessibility, organization, and language equity when used with transparent human verification [2-5]. The integrity problem begins when fluent scholarly expression becomes disconnected from accountable judgment. In health sciences publishing, where claims may influence care, policy, education, research priorities, and public trust, the crucial question is not only whether AI was used but also whether major claims remain traceable to evidence, interpretation, uncertainty, context, and responsible authorship [7,8]. This viewpoint proposes expression-rich but judgment-thin scholarship as a conceptual lens for recognizing that mismatch and domain judgment traceability as a claim-centered governance response. Detection asks whether text can be identified as machine generated; disclosure asks whether AI use has been reported; and judgment traceability asks whether the claims carried by the text can be reconstructed and defended by responsible authors. The framework does not replace disclosure, reporting guidelines, peer review, or research integrity standards. It adds a proportional question at the point where those mechanisms converge: does the strength of a consequential claim remain matched to its evidence, interpretation, limits, and accountable

human judgment? In an era when scholarly expression can be generated at scale, responsible publishing will increasingly depend not only on knowing how text was produced but also on whether the claims it carries are proportionate, defensible, and traceable to accountable human judgment [4,6-9,15].

## Acknowledgments

The authors thank the editors and reviewers for their constructive comments, which strengthened the conceptual precision, evidentiary grounding, and practical orientation of this manuscript.

Using the GAIDeT (Generative Artificial Intelligence Delegation Taxonomy) declaration framework [16], the authors report that generative AI (GAI) tools, including ChatGPT by OpenAI, were used during manuscript development and revision in 2026 to support conceptual discussion, structural editing, manuscript revision, literature search support, summarization support, reference-checking support, language refinement, proofreading, and language polishing. These delegated tasks correspond to GAIDeT categories related to conceptualization, literature review support, writing and editing, and publication support. The authors did not use GAI to replace scholarly, methodological, clinical, or editorial judgment and did not treat GAI output as authoritative evidence.

## Funding

No financial support was received for this work.

## Data Availability

Data sharing is not applicable to this viewpoint because no empirical datasets were generated or analyzed.

## Authors' Contributions

TH conceptualized the manuscript, developed the central argument and proposed framework, drafted the manuscript, revised the intellectual content, verified references and claim-to-source correspondence, and approved the final version for submission. YTC contributed to conceptual refinement, interpretation of the literature, critical revision of the manuscript for important intellectual content, reference verification, and approval of the final version for submission. All substantive claims, interpretations, recommendations, source selection, and final decisions were reviewed and revised by the authors. Final verification of bibliographic details, publication status, DOI information, author names, and claim-to-source correspondence was performed by the authors using publisher pages, indexing records, PubMed records, arXiv or SSRN records, or full-text sources, when available. No confidential patient information, unpublished data, or identifiable personal data were entered into GAI tools. Both authors reviewed the final manuscript, can take responsibility for the central claims and cited sources relevant to their contributions, and accept full responsibility for the final manuscript, including all claims, interpretations, references, and conclusions.

## Conflicts of Interest

None declared.

## References

1. Kobak D, González-Márquez R, Horvát E-Á, Lause J. Delving into LLM-assisted writing in biomedical publications through excess vocabulary. *Sci Adv*. Jul 04, 2025;11(27):eadt3813. [FREE Full text] [doi: [10.1126/sciadv.adt3813](https://doi.org/10.1126/sciadv.adt3813)] [Medline: [40601754](https://pubmed.ncbi.nlm.nih.gov/40601754/)]
2. Amano T, Ramírez-Castañeda V, Berdejo-Espinola V, Borokini I, Chowdhury S, Golivets M, et al. The manifold costs of being a non-native English speaker in science. *PLoS Biol*. Jul 18, 2023;21(7):e3002184. [FREE Full text] [doi: [10.1371/journal.pbio.3002184](https://doi.org/10.1371/journal.pbio.3002184)] [Medline: [37463136](https://pubmed.ncbi.nlm.nih.gov/37463136/)]
3. Tang A, Li K, Kwok KO, Cao L, Luong S, Tam W. The importance of transparency: declaring the use of generative artificial intelligence (AI) in academic writing. *J Nurs Scholarsh*. Mar 2024;56(2):314-318. [doi: [10.1111/jnu.12938](https://doi.org/10.1111/jnu.12938)] [Medline: [37904646](https://pubmed.ncbi.nlm.nih.gov/37904646/)]
4. Cohen JF, Moher D. Generative artificial intelligence and academic writing: friend or foe? *J Clin Epidemiol*. Mar 2025;179:111646. [FREE Full text] [doi: [10.1016/j.jclinepi.2024.111646](https://doi.org/10.1016/j.jclinepi.2024.111646)] [Medline: [39706536](https://pubmed.ncbi.nlm.nih.gov/39706536/)]
5. Bagenal J, Biamis C, Boillot M, Brierley R, Chew M, Dehnel T, et al. Generative AI: ensuring transparency and emphasising human intelligence and accountability. *Lancet*. Nov 30, 2024;404(10468):2142-2143. [FREE Full text] [doi: [10.1016/S0140-6736\(24\)02615-1](https://doi.org/10.1016/S0140-6736(24)02615-1)] [Medline: [39615987](https://pubmed.ncbi.nlm.nih.gov/39615987/)]
6. Messeri L, Crockett MJ. Artificial intelligence and illusions of understanding in scientific research. *Nature*. Mar 06, 2024;627(8002):49-58. [FREE Full text] [doi: [10.1038/s41586-024-07146-0](https://doi.org/10.1038/s41586-024-07146-0)] [Medline: [38448693](https://pubmed.ncbi.nlm.nih.gov/38448693/)]
7. van Zoonen W, Morgan-Thomas A, Tursunbayeva A. Beyond AI disclosure: claim accountability and responsible research in scholarly publishing. *Eur Manag J*. Aug 2026;44(4):550-556. [FREE Full text] [doi: [10.1016/j.emj.2026.06.001](https://doi.org/10.1016/j.emj.2026.06.001)]

8. Johnson RF. Provenance, not prohibition: a framework for AI in scholarly publishing. *Laryngoscope Investig Otolaryngol*. Aug 2026;11(4):e70487. [FREE Full text] [doi: [10.1002/lio2.70487](https://doi.org/10.1002/lio2.70487)] [Medline: [42382923](https://pubmed.ncbi.nlm.nih.gov/42382923/)]
9. Moffatt B, Hall A. Is AI my co-author? The ethics of using artificial intelligence in scientific publishing. *Account Res*. Nov 2025;32(8):1313-1329. [doi: [10.1080/08989621.2024.2386285](https://doi.org/10.1080/08989621.2024.2386285)] [Medline: [39109816](https://pubmed.ncbi.nlm.nih.gov/39109816/)]
10. Cleland J, Driessen E, Masters K, Lingard L, Maggio LA. When and how to disclose AI use in academic publishing: AMEE guide no.192. *Med Teach*. Dec 30, 2025;48(4):542-553. [FREE Full text] [doi: [10.1080/0142159x.2025.2607513](https://doi.org/10.1080/0142159x.2025.2607513)]
11. Zielinski C, Winker MA, Aggarwal R, Ferris LE, Heinemann M, Lapeña JF, et al. Chatbots, generative AI, and scholarly manuscripts: WAME recommendations on chatbots and generative artificial intelligence in relation to scholarly publications. *Colomb Med (Cali)*. Dec 01, 2023;54(3):e1015868. [FREE Full text] [doi: [10.25100/cm.v54i3.5868](https://doi.org/10.25100/cm.v54i3.5868)] [Medline: [38089825](https://pubmed.ncbi.nlm.nih.gov/38089825/)]
12. Flanagan A, Pirracchio R, Khera R, Berkwits M, Hswen Y, Bibbins-Domingo K. Reporting use of AI in research and scholarly publication—JAMA Network guidance. *JAMA*. Apr 02, 2024;331(13):1096-1098. [FREE Full text] [doi: [10.1001/jama.2024.3471](https://doi.org/10.1001/jama.2024.3471)] [Medline: [38451540](https://pubmed.ncbi.nlm.nih.gov/38451540/)]
13. Ganjavi C, Eppler MB, Pekcan A, Biedermann B, Abreu A, Collins GS, et al. Publishers' and journals' instructions to authors on use of generative artificial intelligence in academic and scientific publishing: bibliometric analysis. *BMJ*. Jan 31, 2024;384:e077192. [FREE Full text] [doi: [10.1136/bmj-2023-077192](https://doi.org/10.1136/bmj-2023-077192)] [Medline: [38296328](https://pubmed.ncbi.nlm.nih.gov/38296328/)]
14. Yoo JH. Defining the boundaries of AI use in scientific writing: a comparative review of editorial policies. *J Korean Med Sci*. Jun 16, 2025;40(23):e187. [FREE Full text] [doi: [10.3346/jkms.2025.40.e187](https://doi.org/10.3346/jkms.2025.40.e187)] [Medline: [40524628](https://pubmed.ncbi.nlm.nih.gov/40524628/)]
15. Park J. Towards a transparent and reproducible AI-assisted research paper writing. *Genomics Inform*. Dec 02, 2025;23(1):26. [FREE Full text] [doi: [10.1186/s44342-025-00057-0](https://doi.org/10.1186/s44342-025-00057-0)] [Medline: [41331701](https://pubmed.ncbi.nlm.nih.gov/41331701/)]
16. Suchikova Y, Tsybuliak N, Teixeira da Silva JA, Nazarovets S. GAIDeT (Generative AI Delegation Taxonomy): a taxonomy for humans to delegate tasks to generative artificial intelligence in scientific research and publishing. *Account Res*. Apr 2026;33(3):2544331. [FREE Full text] [doi: [10.1080/08989621.2025.2544331](https://doi.org/10.1080/08989621.2025.2544331)] [Medline: [40781729](https://pubmed.ncbi.nlm.nih.gov/40781729/)]
17. Májovský M, Černý M, Kasal M, Komarc M, Netuka D. Artificial intelligence can generate fraudulent but authentic-looking scientific medical articles: Pandora's box has been opened. *J Med Internet Res*. May 31, 2023;25:e46924. [FREE Full text] [doi: [10.2196/46924](https://doi.org/10.2196/46924)] [Medline: [37256685](https://pubmed.ncbi.nlm.nih.gov/37256685/)]
18. Walters WH, Wilder EI. Fabrication and errors in the bibliographic citations generated by ChatGPT. *Sci Rep*. Sep 07, 2023;13(1):14045. [FREE Full text] [doi: [10.1038/s41598-023-41032-5](https://doi.org/10.1038/s41598-023-41032-5)] [Medline: [37679503](https://pubmed.ncbi.nlm.nih.gov/37679503/)]
19. Resnik DB, Hosseini M. Hallucinated citations produced by generative artificial intelligence may constitute research misconduct when citations function as data in scholarly papers. *Account Res*. Oct 15, 2026;33(7):2645390. [FREE Full text] [doi: [10.1080/08989621.2026.2645390](https://doi.org/10.1080/08989621.2026.2645390)] [Medline: [41833014](https://pubmed.ncbi.nlm.nih.gov/41833014/)]
20. Bender EM, Gebru T, McMillan-Major A, Shmitchell S. On the dangers of stochastic parrots: can language models be too big? New York, NY. Association for Computing Machinery; 2021. Presented at: 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21); March 3-10, 2021:610-623; Canada. [doi: [10.1145/3442188.3445922](https://doi.org/10.1145/3442188.3445922)]
21. Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. *J Am Med Inform Assoc*. 2012;19(1):121-127. [FREE Full text] [doi: [10.1136/amiajnl-2011-000089](https://doi.org/10.1136/amiajnl-2011-000089)] [Medline: [21685142](https://pubmed.ncbi.nlm.nih.gov/21685142/)]
22. Lyell D, Coiera E. Automation bias and verification complexity: a systematic review. *J Am Med Inform Assoc*. Mar 01, 2017;24(2):423-431. [FREE Full text] [doi: [10.1093/jamia/ocw105](https://doi.org/10.1093/jamia/ocw105)] [Medline: [27516495](https://pubmed.ncbi.nlm.nih.gov/27516495/)]
23. Burrell J. How the machine 'thinks': understanding opacity in machine learning algorithms. *Big Data Soc*. Jan 06, 2016;3(1):1-12. [FREE Full text] [doi: [10.1177/2053951715622512](https://doi.org/10.1177/2053951715622512)]
24. Parker L, Boughton S, Bero L, Byrne JA. Paper mill challenges: past, present, and future. *J Clin Epidemiol*. Dec 2024;176:111549. [FREE Full text] [doi: [10.1016/j.jclinepi.2024.111549](https://doi.org/10.1016/j.jclinepi.2024.111549)] [Medline: [39389277](https://pubmed.ncbi.nlm.nih.gov/39389277/)]
25. Candal-Pedreira C, Ross JS, Ruano-Ravina A, Egilman DS, Fernández E, Pérez-Ríos M. Retracted papers originating from paper mills: cross sectional study. *BMJ*. Nov 28, 2022;379:e071517. [FREE Full text] [doi: [10.1136/bmj-2022-071517](https://doi.org/10.1136/bmj-2022-071517)] [Medline: [36442874](https://pubmed.ncbi.nlm.nih.gov/36442874/)]
26. Liang W, Yuksekgonul M, Mao Y, Wu E, Zou J. GPT detectors are biased against non-native English writers. *Patterns (N Y)*. Jul 14, 2023;4(7):100779. [FREE Full text] [doi: [10.1016/j.patter.2023.100779](https://doi.org/10.1016/j.patter.2023.100779)] [Medline: [37521038](https://pubmed.ncbi.nlm.nih.gov/37521038/)]
27. Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. *Lancet*. Jul 04, 2009;374(9683):86-89. [doi: [10.1016/S0140-6736\(09\)60329-9](https://doi.org/10.1016/S0140-6736(09)60329-9)] [Medline: [19525005](https://pubmed.ncbi.nlm.nih.gov/19525005/)]
28. Macleod MR, Michie S, Roberts I, Dirnagl U, Chalmers I, Ioannidis JPA, et al. Biomedical research: increasing value, reducing waste. *Lancet*. Jan 11, 2014;383(9912):101-104. [doi: [10.1016/S0140-6736\(13\)62329-6](https://doi.org/10.1016/S0140-6736(13)62329-6)] [Medline: [24411643](https://pubmed.ncbi.nlm.nih.gov/24411643/)]
29. Moher D, Glasziou P, Chalmers I, Nasser M, Bossuyt PMM, Korevaar DA, et al. Increasing value and reducing waste in biomedical research: who's listening? *Lancet*. Apr 09, 2016;387(10027):1573-1586. [doi: [10.1016/S0140-6736\(15\)00307-4](https://doi.org/10.1016/S0140-6736(15)00307-4)] [Medline: [26423180](https://pubmed.ncbi.nlm.nih.gov/26423180/)]
30. Tang G, Cai H. Citation contamination by paper mill articles in systematic reviews of the life sciences. *JAMA Netw Open*. Jun 02, 2025;8(6):e2515160. [FREE Full text] [doi: [10.1001/jamanetworkopen.2025.15160](https://doi.org/10.1001/jamanetworkopen.2025.15160)] [Medline: [40504528](https://pubmed.ncbi.nlm.nih.gov/40504528/)]
31. Tomkins A, Zhang M, Heavlin WD. Reviewer bias in single- versus double-blind peer review. *Proc Natl Acad Sci U S A*. Nov 28, 2017;114(48):12708-12713. [FREE Full text] [doi: [10.1073/pnas.1707323114](https://doi.org/10.1073/pnas.1707323114)] [Medline: [29138317](https://pubmed.ncbi.nlm.nih.gov/29138317/)]
32. Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ*. Jan 13, 1996;312(7023):71-72. [FREE Full text] [doi: [10.1136/bmj.312.7023.71](https://doi.org/10.1136/bmj.312.7023.71)] [Medline: [8555924](https://pubmed.ncbi.nlm.nih.gov/8555924/)]

33. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK, SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nat Med*. Sep 09, 2020;26(9):1364-1374. [FREE Full text] [doi: [10.1038/s41591-020-1034-x](https://doi.org/10.1038/s41591-020-1034-x)] [Medline: [32908283](https://pubmed.ncbi.nlm.nih.gov/32908283/)]
34. Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, et al. DECIDE-AI expert group. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *Nat Med*. May 18, 2022;28(5):924-933. [FREE Full text] [doi: [10.1038/s41591-022-01772-9](https://doi.org/10.1038/s41591-022-01772-9)] [Medline: [35585198](https://pubmed.ncbi.nlm.nih.gov/35585198/)]
35. Kim HJ, Ko RE, Lim SY, Park S, Suh GY, Lee YJ. Sepsis alert systems, mortality, and adherence in emergency departments: a systematic review and meta-analysis. *JAMA Netw Open*. Jul 01, 2024;7(7):e2422823. [FREE Full text] [doi: [10.1001/jamanetworkopen.2024.22823](https://doi.org/10.1001/jamanetworkopen.2024.22823)] [Medline: [39037814](https://pubmed.ncbi.nlm.nih.gov/39037814/)]
36. Wang H, Li Y, Naidech A, Luo Y. Comparison between machine learning methods for mortality prediction for sepsis patients with different social determinants. *BMC Med Inform Decis Mak*. Jun 16, 2022;22(Suppl 2):156. [FREE Full text] [doi: [10.1186/s12911-022-01871-0](https://doi.org/10.1186/s12911-022-01871-0)] [Medline: [35710407](https://pubmed.ncbi.nlm.nih.gov/35710407/)]
37. Hwang MI, Bond WF, Powell ES. Sepsis alerts in emergency departments: a systematic review of accuracy and quality measure impact. *West J Emerg Med*. Aug 24, 2020;21(5):1201-1210. [FREE Full text] [doi: [10.5811/westjem.2020.5.46010](https://doi.org/10.5811/westjem.2020.5.46010)] [Medline: [32970576](https://pubmed.ncbi.nlm.nih.gov/32970576/)]

## Abbreviations

**AI:** artificial intelligence

**CONSORT-AI:** Consolidated Standards of Reporting Trials–Artificial Intelligence

**DECIDE-AI:** Developmental and Exploratory Clinical Investigation of Decision Support Systems Driven by Artificial Intelligence

**ED:** emergency department

**GAI:** generative AI

**GAIDeT:** Generative Artificial Intelligence Delegation Taxonomy

**LLM:** large language model

*Edited by S Brini; submitted 02.Jul.2026; peer-reviewed by D Barrett, U Omar; comments to author 22.Jul.2026; revised version received 26.Sep.2026; accepted 02.Oct.2026; published 19.Oct.2026*

*Please cite as:*

Hsieh T, Chen Y-T

*From AI Disclosure to Judgment Traceability: Addressing Research Waste in AI-Assisted Health Sciences Publishing*

*J Meta and Res Int* 2026;1:e106068

URL: <https://metascience.jmir.org/2026/1/e106068>

doi: [10.2196/106068](https://doi.org/10.2196/106068)

PMID:

©Tino Hsieh, Ya-Ting Chen. Originally published in the JMIR Metascience and Research Integrity (<https://metascience.jmir.org>), 19.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Metascience and Research Integrity, is properly cited. The complete bibliographic information, a link to the original publication on <https://metascience.jmir.org>, as well as this copyright and license information must be included.