

Viewpoint

Derivative Epistemic Agency: AI-Mediated Scientific Cognition and the Limits of Human Answerability

Åke Elden, PhD

Department of Research, NLA University College, Oslo, Oslo, Norway

Corresponding Author:

Åke Elden, PhD

Department of Research

NLA University College

Pilestredet park 33

Oslo, Oslo, 0167

Norway

Phone: 47 45312255

Email: Ake.Elden@nla.no

Abstract

Generative AI can assist in scientific work, while leaving authors responsible for claims they may not adequately understand. Existing authorship and disclosure frameworks establish human accountability, but declaring supervision does not demonstrate the competence required to exercise it. This conceptual viewpoint examines when reliance on generative AI becomes incompatible with scientific answerability and proposes a proportionate response for biomedical publishing. The analysis distinguishes integrated epistemic extension from derivative epistemic substitution along a graded continuum of origination, comprehension, evaluation, substantive answerability, control, provenance awareness, and competence retention. Origination alone is not disqualifying: researchers may responsibly adopt ideas suggested by systems or collaborators. The central requirement is sufficient justificatory competence, assessed through the ability to address reasonable specialist challenges concerning evidence, assumptions, limitations, alternatives, and possible revision. This requirement applies across the author team and scales with the importance of the claim and the consequences of error. It does not require reproducing every cognitive operation, explaining model internals, or abandoning justified reliance on independently established reliability. Responsibility asymmetry arises when authors retain formal responsibility, AI supplies substantial inferential content, and the team lacks adequate coverage of the capacities needed to justify adopting that content. The problem is, therefore, not the proportion of generated text. A single unsupported inference can matter more than extensive language assistance. Work on distributed cognition, scientific understanding, automation, and research integrity supports background analysis. Evidence concerning cognitive offloading and possible deskilling is discussed with explicit limits: observational clinical findings and self-reported knowledge work cannot establish that generative AI causes lasting competence loss among scientists. Institutional incentives are treated as a possible reinforcing mechanism rather than an established causal explanation. The viewpoint compares authorship criteria, contributor roles, methodological attribution, and recent AI reporting frameworks, including GAIDeT (Generative Artificial Intelligence Delegation Taxonomy) and GAMER (Generative Artificial Intelligence Tools in Medical Research). These instruments offer complementary forms of transparency and verification, but reporting activities does not itself establish retained capacity to defend a consequential inference. Proposed extensions include an epistemic participation profile linking delegated operations to named authors; reporting at the manuscript, operation, or claim level according to risk; and targeted reasoning audits triggered by unresolved concerns. Audits should assess substance; permit language assistance and accommodations; and provide confidentiality, opportunities to respond, and appeal. Remediation should prioritize restoring or supplementing competence, narrowing claims, or declining to adopt unsupported conclusions. These instruments are normative prototypes, not validated measures or established editorial requirements. Pilots should evaluate feasibility, agreement between assessors, erroneous concerns, successful remediation, and unequal effects across researcher groups. The framework preserves legitimate collaboration and linguistic assistance while identifying a question that disclosure alone cannot settle: whether human authors can justify the scientific claims they adopt. Its contribution is an operational specification of answerability rather than a general objection to technologically distributed cognition.

(*J Meta and Res Int* 2026;1:e105907) doi: [10.2196/105907](https://doi.org/10.2196/105907)

KEYWORDS

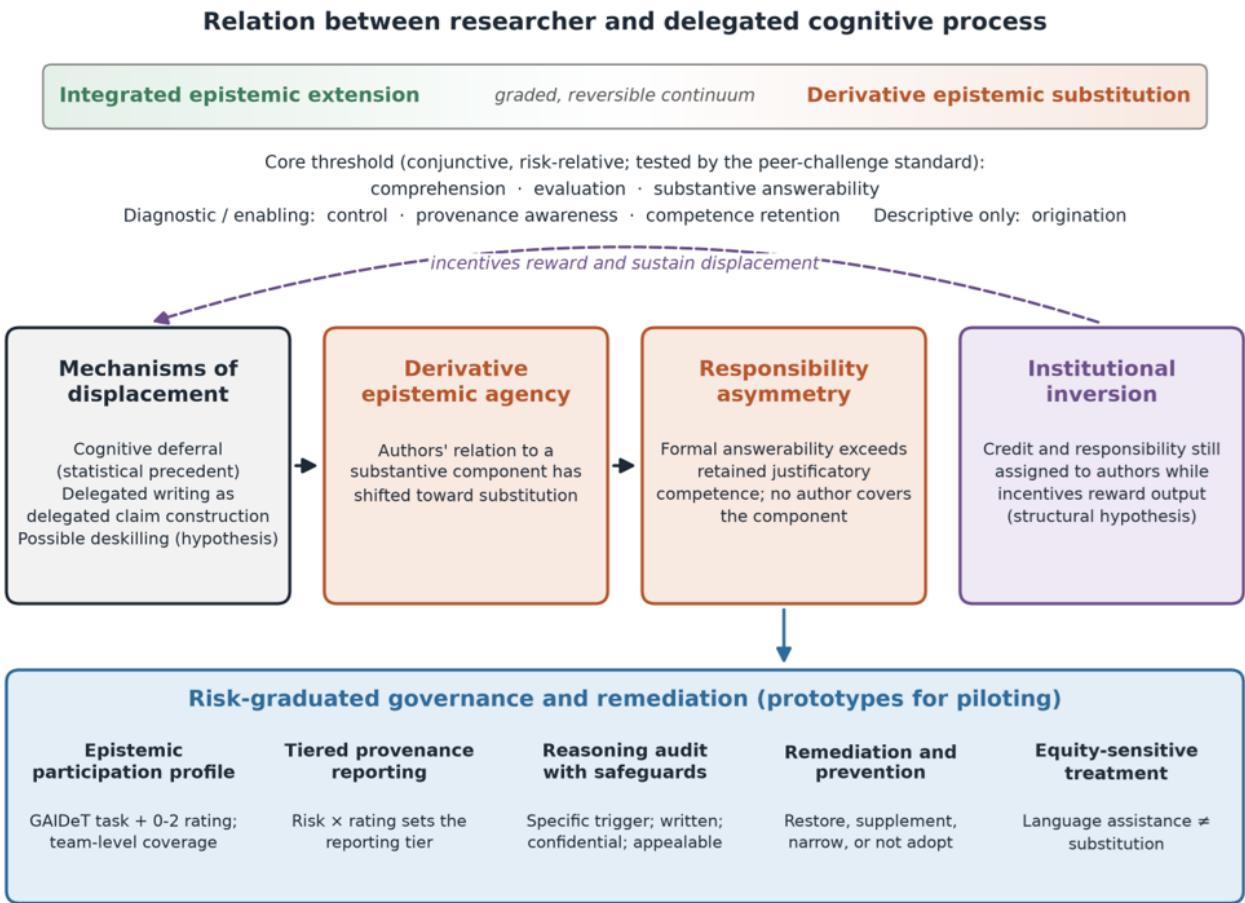
artificial intelligence; AI; scientific authorship; research integrity; epistemic responsibility; distributed cognition; responsibility gap; contributorship; metascience; governance frameworks; generative AI disclosure

Background

Large language models (LLMs) are increasingly used in scientific writing. An analysis of more than 15 million PubMed abstracts estimated that at least 13.5% of abstracts published in 2024 had been processed with LLMs [1]. A separate analysis of more than one million papers and preprints found increasing estimated LLM modification across several research fields [2]. These are population estimates inferred from linguistic patterns; they do not establish individual authors’ practices or understanding. The International Committee of Medical Journal Editors (ICMJE) and the Committee on Publication Ethics (COPE) require disclosure of AI use and retain responsibility with human authors [3,4].

This viewpoint asks whether those authors retain the capacities needed to discharge that responsibility. Concern about an apparent increase in productivity accompanied by reduced understanding has already been articulated by Messeri and Crockett [5]. The argument in this viewpoint concerns a narrower relation: whether the author team can justify adopting the inferential content of a particular scientific claim. It does not equate assistance with dependence, dependence with incompetence, or an erroneous result with irresponsible conduct. The proposed contribution is a risk-relative specification of answerability and a set of editorial prototypes for evaluating it, rather than the discovery of a wholly new responsibility problem. Figure 1 summarizes the framework and its proposed mechanisms [5-7].

Figure 1. Conceptual framework of the viewpoint. The continuum, definitions, and governance response are the author’s proposals. Arrows indicate hypothesized relations, not demonstrated causal effects. The framework is informed by work on scientific understanding, meaningful human control, and institutional incentives [5-7]. Team-level coverage includes the capacity to justify reliance on independently validated tools; it does not require access to model internals. GAIDeT: Generative Artificial Intelligence Delegation Taxonomy.



Approach and Scope

This viewpoint uses conceptual analysis and a targeted, nonsystematic examination of the literature on scientific understanding, distributed cognition, automation, and research integrity. It compares ICMJE and COPE guidance [3,4,8],

CRedit (Contributor Roles Taxonomy) and MeRiT (Method Reporting with Initials for Transparency) contribution reporting [9,10], and GAIDeT (Generative Artificial Intelligence Delegation Taxonomy) and GAMER (Generative Artificial Intelligence Tools in Medical Research) reporting of AI use [11,12]. These instruments were selected for their relevance to

authorship, delegation, and verification, without claiming that they exhaust current governance practice. The analysis develops definitions, tests them through explicitly hypothetical cases, and derives proposals for evaluation. Its scope is generative AI used to produce, interpret, or write up scientific claims, particularly in biomedical publishing; validated prediction instruments enter only as boundary cases. It does not establish prevalence, causal effects, legal liability, or the effectiveness of the proposed instruments. Coverage is selective and may omit relevant literature, including work outside English-language biomedical publishing and philosophy of science.

Epistemic Labor as a Distributed, Not Purely Individual, Achievement

Scientific knowledge depends on collaboration and instruments whose workings no individual scientist reproduces completely. Recent accounts of distributed cognition emphasize the integration of people and artifacts, while current social epistemology examines how researchers can reasonably rely on opaque systems [13,14]. Consequently, the argument cannot begin from a supposed historical requirement that each author personally perform every intellectual operation. The normative requirement defended here is instead sufficient integration into the process: authors must understand the role of the evidence and be able to justify the grounds on which they adopt it. This is compatible with the ICMJE's division of responsibilities among coauthors [8]. Distribution is therefore a starting condition of the analysis, not evidence of a deficit.

AI assistance raises a particular allocation problem because the system cannot assume the authorship responsibilities that publishing guidance assigns to people [3,4]. That fact does not entail that reliance on a system is unjustified. Ortmann [14] argues that empirically established reliability, together with appropriate reliance on the researchers who evaluated it, can justify using an opaque scientific tool. In the framework in this viewpoint, justified reliance on an independently validated computational pipeline is not derivative substitution. Authors need not explain a model's internal computation, when they can justify the relevance and adequacy of its independent validation for the claim they make. The problematic case is adoption of a substantive inference for which the team can supply neither an adequate justification nor a defensible account of validated reliability.

The difference from human collaboration must therefore be stated carefully. A human collaborator's mere availability does not guarantee competence, and a system's inability to bear responsibility does not make its outputs unusable. What matters is whether the author team can answer for the contribution or for its justified use [8,14]. For example, a statistician may defend an analysis that a clinician cannot independently reproduce; a generated interpretation accepted by both without understanding

has no corresponding coverage. This comparison is an analytical case, not evidence that all human collaboration succeeds or that all AI use fails. Responsibility asymmetry names the uncovered justificatory obligation, while the concern about fluent output obscuring limited understanding supplies one plausible route to it [5].

Extended Cognition and a Continuum of Epistemic Relations

Distributed cognition provides an important objection to any demand for unaided authorship: artifacts can participate in cognitive processes rather than merely transmit completed thoughts [13]. The distinction proposed here therefore concerns integration, not the physical location of cognition. A researcher who uses computational support while understanding its evidential role may remain responsible for the resulting judgment; a researcher who accepts a conclusion without a defensible basis may not meet that standard. This is a normative interpretation of distributed responsibility, not a finding that external cognitive support reduces agency. The recent debate over justified reliance on opaque AI reinforces the need to distinguish understanding a tool internally from understanding why its output is acceptable in a specified context [14].

Delegated writing deserves particular attention because drafting may change what is claimed. Bietti and Bangerter [15] argue that scientific writing is itself an epistemic practice and that routine outsourcing could separate textual production from critical thinking. This is a conceptual warning, not proof of inevitable deterioration. The relevant implication for this viewpoint is conditional: if a draft introduces a causal explanation, an inferential bridge, or an interpretation absent from the author's notes, adopting it requires evaluation of that addition. Language improvement that preserves already defensible content need not have the same significance. Thus, the description "writing assistance" is insufficient to classify the epistemic relation; the added content and the authors' grounds for adopting it must be examined.

The proposed continuum has seven dimensions: origination, comprehension, evaluation, substantive answerability, control, provenance awareness, and competence retention. It draws on work distinguishing cognitive integration, meaningful human control, and scientific understanding [6,13,16,17], but the selection and grouping of these dimensions are this paper's analytical proposal. Table 1 sets out their roles and the two poles, integrated epistemic extension and derivative epistemic substitution. The dimensions are not a validated psychometric scale, and the table does not imply that they can be added into a single numerical measure. The next section specifies which capacities are central to the proposed threshold and which serve only as diagnostic indicators.

Table 1. Proposed dimensions of epistemic integration and substitution, informed by distributed cognition, meaningful human control, and scientific understanding (the taxonomy and threshold roles are the author's proposal) [6,13,16,17].

Dimension	Pole of integrated extension	Pole of derivative substitution	Analytical level; role in threshold
Origination	The researcher generates the candidate inferential structure.	The system generates the candidate inferential structure.	Causal history of the content; descriptive only, not a threshold condition
Comprehension	The researcher understands the grounds and implications of the content adopted.	The researcher cannot explain the grounds of the content adopted.	Capacity of the author relative to a claim; core and necessary
Evaluation	The researcher can assess validity and evidential support through an epistemically independent route.	The researcher can only assess plausibility, not validity.	Capacity of the author relative to a claim; core and necessary
Substantive answerability	The researcher can defend and, if needed, revise the adopted claim under challenge.	The researcher depends on the system to supply a justification they cannot evaluate.	Capacity of the author relative to a claim; core and necessary
Control	The researcher can modify, reject, or redirect the process.	The researcher can only accept or discard the output as given.	Author's relation to the production process; enabling or diagnostic
Provenance awareness	The researcher knows which elements originated with the system.	The origin of specific inferential elements is no longer tracked.	Second-order knowledge about the content; diagnostic
Competence retention	The researcher could detect a significant error without relying on a generative system.	Error detection itself depends on a generative system.	Dispositional capacity over time; diagnostic

The dimensions operate at different levels and are not normatively equivalent. Origination describes the causal history of content and is not independently disqualifying; adoption can be justified even when an idea originated with a collaborator or system. Control, provenance awareness, and competence retention are enabling or diagnostic considerations; their absence can warrant closer examination without automatically establishing a failure of answerability. Comprehension, evaluation, and substantive answerability are the three core capacities in this framework, informed by accounts of understanding and responsibility [6,16,17]. A researcher may understand an argument without having originated it or responsibly accept an instrument's output on the basis of independent validation without reconstructing its internal computation [14]. Sufficient justificatory competence therefore concerns the grounds for accepting the claim, including justified reliance, rather than unaided reproduction of every operation.

A Threshold for Sufficient Justificatory Competence

The proposed peer-challenge standard gives this requirement a practical form: authors should be able to address the challenges a competent specialist referee could reasonably raise about a claim's grounds, assumptions, limitations, alternatives, and revision conditions. Comprehension concerns what the claim means and depends on; evaluation concerns whether the evidence supports it; and substantive answerability concerns defending, qualifying, or withdrawing it under challenge. This operational proposal draws on accounts of scientific understanding and competent verification [16-18] rather than claiming to reproduce an existing validated assessment. Authors may consult sources, records, code, collaborators, and language assistance. The test fails only when they must rely on generative AI to supply a justification they cannot themselves evaluate; it is neither a memory examination nor a ban on external support.

The standard is conjunctive across the core capacities and risk-relative in depth. A peripheral background claim calls for less extensive justification than an interpretation of a primary outcome with clinical implications. This proportionality extends the emphasis on competent, appropriately scoped verification in recent authorship ethics [18]. Routine attestation should support an initial presumption of adequate competence. A specific concern about a consequential inference, rather than AI use or an unfamiliar writing style, is needed to justify additional scrutiny. Failure on one relevant capacity cannot be cancelled by strength on an unrelated dimension.

The assessment applies to the author team rather than requiring every author to master every component. The ICMJE expects authors to identify coauthors responsible for other parts of the work and to have confidence in those contributions [8]. The coverage requirement proposed here makes that expectation more explicit: for each substantive component, a named author or an identified combination of authors must be able to justify the contribution or the warranted reliance on it. The team may rely on outside scholarship and validated tools; the requirement is not that their originators become coauthors [14]. For an AI-generated interpretation, however, knowing which system produced it is not equivalent to identifying a defensible basis for accepting it. The uncovered case is one in which the team can neither defend the interpretation nor establish adequate grounds for relying on it.

Three terms are used consistently throughout the analysis: "Derivative epistemic agency" denotes an author's relation to substantive content when that relation approaches substitution; "responsibility asymmetry" denotes the mismatch between formal responsibility and inadequate team-level coverage; and "institutional inversion" denotes a proposed incentive mechanism that could sustain this mismatch. These are analytical definitions developed here, situated in existing discussions of understanding, control, and responsibility [5,6,18]. Table 2 tests their application through six hypothetical cases, including

specialist collaboration, an AI-originated hypothesis, and drafting that introduces an unsupported causal interpretation.

The cases illustrate the framework’s commitments and limits; they are not observations of actual researchers.

Table 2. Six hypothetical cases applying the proposed threshold.^a

Case	Analysis	Classification and consequence
(A) Language editing. A researcher writing in a second language drafts a discussion section and uses an LLM ^b to improve grammar and fluency; the claims and their structure are unchanged.	Presentational content only; all core capacities retained.	Integrated extension. Presentational only, so outside the participation profile; manuscript-level disclosure.
(B) Specialist human collaborator. A clinician coauthor cannot evaluate the mixed-effects model on which the primary outcome rests, but knows that the statistician coauthor specified it, can state its role in the argument, and relies on that coauthor’s expertise.	The statistician meets the peer-challenge standard; the clinician’s reliance is rational trust in testimony.	Legitimate reliance; team-level coverage met. Ordinary contribution reporting applies; no additional AI report is required for this wholly human case.
(C) AI in place of the statistician. The same model is specified and interpreted with an LLM; no author can explain the choice of random-effects structure or how its assumptions were checked.	Central, high-risk claim; comprehension and evaluation are absent for a substantial component that no author covers.	Derivative substitution and responsibility asymmetry. Reporting does not cure the deficit: competence must be restored (eg, a statistician coauthor who reanalyzes the data), the claim narrowed, or the claim withdrawn.
(D) AI-originated hypothesis. An LLM suggests a hypothesis the authors had not considered; they understand its rationale, design and run the test themselves, and can defend the design.	System origination only; core capacities retained.	Integrated extension if evaluation and answerability are retained. Profile entry; risk follows the importance of dependent claims, not AI origination alone.
(E) AI-assisted literature synthesis (boundary case). An LLM-based tool screens and summarizes 200 studies for a narrative review; the authors verify a sample of summaries and can explain the structure of the synthesis but cannot vouch for unverified summaries.	Competence is partial; its adequacy depends on how far the conclusions rest on the unverified summaries.	Sampling is not automatically deficient. Authors must justify its adequacy for the synthesis. If unverified content leaves a consequential inference without adequate support, further verification or narrower conclusions are required; a high-risk operation requires rating 2.
(F) “Only writing” (boundary case). An author asks an LLM to turn notes on results into a discussion; the draft introduces a causal interpretation absent from the notes, which the author approves without examining whether the design supports it.	Delegated drafting has become delegated claim construction; evaluation is absent for the new claim.	Substitution for that claim, although AI was used “only for writing.” The claim must be justified, qualified, or removed.

^aThe cases are analytical illustrations, not empirical findings; legitimate reliance is interpreted in light of authorship guidance and the reliability objection [8,14].

^bLLM: large language model.

*From Extension to Substitution:
Functional Content and Answerable
Judgment*

Functional production of inferential content should be distinguished from undertaking scientific responsibility. An LLM can produce a passage with premises, conclusions, and apparent reasons, but current authorship guidance does not recognize it as a party capable of assuming responsibility for the work [3,4]. This institutional distinction is sufficient for the argument in this viewpoint; it does not require settling whether AI systems understand or reason in every philosophical sense. The question is what human authors must do when adopting such content. They must be able to evaluate its role in their argument and respond to challenges to its adequacy rather than treating the generation of a coherent explanation as evidence that its adoption has been justified. This concern is consistent with recent discussion of the epistemic risks of LLMs in science [19].

*Mechanisms of Displacement: Cognitive
Deferral, Offloading, and Deskilling*

Research on automation supplies candidate mechanisms for displaced evaluation. In an experimental study of electronic prescribing, Lyell et al [20] investigated errors associated with reliance on clinical decision support. Such work supports the concern about automation bias, but its tasks and systems differ from generative drafting or interpretation in research. The relevant extrapolation is therefore a hypothesis: researchers may accept a system’s recommendation, while reducing independent scrutiny. It remains possible that experienced researchers instead use AI to expose alternatives and improve verification. The distinction between helpful support and inadequate checking must be assessed through the particular task and the evidence of evaluation, not inferred from tool use alone.

Statistical interpretation offers an analogous problem. Amrhein et al [21] criticize treating statistical significance as a mechanical dividing line and argue for interpretation attentive to estimates, uncertainty, and context. The analogy is not that statistical software and LLMs are identical but that obtaining a formally



presented output does not settle its scientific meaning. A defensible remedy is to ensure that someone in the author team can justify the analytical choices and interpretation, consistent with current contribution and responsibility guidance [8,22]. The proposed framework extends this logic to AI-assisted inferential content. A fluent account of a model is insufficient if no author can explain why that model addresses the research question or how its assumptions affect the conclusion.

Offloading itself can be adaptive. Gilbert et al's [23] review of intention offloading describes effective uses of external reminders and the metacognitive factors influencing them. Remembering delayed intentions is not equivalent to constructing scientific arguments, so these findings do not establish effects of generative AI on scientific competence. They do show why externalization cannot simply be classified as a loss. In the framework in this viewpoint, assistance is evaluated by the author's retained ability to understand and assess the adopted content. The further suggestion that repeated outsourcing of interpretation might erode competence is a longitudinal hypothesis requiring direct tests in research settings, not a consequence that follows from the definition of offloading.

Available findings warrant caution about both overreliance and overgeneralization. In a multicenter observational study, adenoma detection during standard colonoscopy fell from 28.4% before exposure to 22.4% after exposure to AI-assisted detection [24]. The design does not establish that AI caused permanent deskilling; its clinical findings cannot establish causal or lasting deskilling in scientific writing. In a survey of 319 knowledge workers, greater confidence in generative AI was associated with less self-reported critical thinking [25]. Self-report and association do not demonstrate lasting competence loss or determine its direction of causation. These studies motivate direct investigations, such as repeated evaluations of researchers' reasoning under different assistance conditions; they do not establish the prevalence of derivative epistemic agency.

Responsibility Asymmetry: Formal Conditions

Responsibility asymmetry belongs within an existing family of problems rather than replacing them. Elish's moral crumple zones concern human operators bearing responsibility within systems they only partly control [26]. Rubel et al [27] describe agency laundering, in which technological mediation obscures an agent's responsibility. Meaningful human control connects system behavior with relevant human reasons and identifiable human understanding [6]. These accounts motivate attention to the relationship between responsibility and capacity. The specification in this viewpoint concerns scientific justification: an author may remain the person expected to answer for a claim, while lacking an adequate basis for doing so. It neither establishes that nobody is morally responsible nor absolves a researcher who negligently adopted content.

Under the proposed definition, responsibility asymmetry arises when three conditions coincide: Human authors are formally required to answer for a claim; AI has supplied substantial inferential content adopted in it; and the team lacks adequate

coverage of at least one core capacity needed to justify that adoption at the claim's level of risk. This formulation combines the authorship obligation with the paper's proposed competence threshold [8,18]. Extensive AI use need not satisfy the third condition, while a single pivotal interpretation can do so. The criterion concerns the author team's justificatory relation to the content, not the number of generated words or the mere presence of an error. A correct conclusion can lack adequate justification, and a responsibly justified conclusion can later prove wrong.

The closest recent account is Cassimiro Nascimento and da Silva's [18] analysis of epistemic duty in AI-assisted writing. They distinguish dimensions of responsibility and emphasize competent, proportionate, independent verification. This proposal should be read as an operational extension of that concern, not a demonstration that their account cannot accommodate answerability. A checklist recording verification may leave unresolved whether an author can respond to a substantive challenge; conversely, good verification may already supply that capacity. The added question is how to identify the author coverage, risk level, and explanatory response needed in a particular case. No claim is made that these proposed instruments are uniquely capable of answering it.

The framework also differs in level from broader diagnoses of AI-mediated science. Messeri and Crockett [5] discuss illusions of understanding, while Birhane et al [19] identify risks for scientific practice and trust. Those arguments provide reasons to examine how content is adopted but do not measure the competence of particular author teams. Responsibility asymmetry is offered as one possible local manifestation, not an established explanation of every field-level effect. Its practical usefulness would depend on distinguishing deficient justification from legitimate dependence without treating ordinary uncertainty as misconduct. That is why the proposed instruments include explicit boundary cases, proportionate scrutiny, and opportunities for remediation.

Institutional Inversion: Incentive Structures and the Production Economy of Science

Institutional incentives may amplify these risks. Edwards and Roy [7] argue that hypercompetition and quantitative measures can create pressures adverse to scientific integrity. Hanson et al [28] document substantial growth in publication volume and the associated strain on publishing. Neither source establishes an AI-induced loss of authors' competence. Together, however, they motivate a structural hypothesis: incentives favoring volume and speed could make inadequately evaluated delegation attractive. This hypothesis concerns an interaction between incentives, tool use, and verification practices; it should not be treated as a general explanation of researchers' motives or as evidence that productivity and understanding must conflict.

"Institutional inversion" is the name proposed here for a specific version of that interaction: institutions continue to attribute epistemic achievement and responsibility to authors, while rewarding arrangements that leave them unable to justify the adopted content. This develops the incentive concern in the

preceding literature [7,28] into a hypothesis about the distribution of cognitive labor. Efficient assistance would not constitute inversion when authors retain adequate understanding and evaluation. Nor does greater output alone establish it. The hypothesis requires evidence that a rewarded production arrangement systematically weakens justificatory coverage, while the corresponding attribution of achievement remains in place.

Corpus studies indicate substantial LLM involvement in scientific writing [1,2], and publishing research documents increasing output pressures [28]. Their coexistence does not establish the proposed causal link. A suitable study would connect task-specific AI use with independent assessment of the team’s ability to justify consequential claims, and then examine whether that relationship varies with assessment regimes, career incentives, or available support. Longitudinal or carefully controlled designs would help separate prior competence, selection into AI use, and changes following its

adoption. Until such evidence is available, institutional inversion remains a mechanism proposed for investigation.

Where Existing Governance Frameworks Fall Short

Existing frameworks address different parts of this problem. McNutt et al [22] discuss transparent attribution of contributions and responsibilities, CRediT specifies contributor roles [9], and MeRiT links methodological activities to named contributors [10]. The ICMJE and COPE allocate authorship responsibilities and address AI use [3,4,8], while GAIDeT and GAMER provide more detailed reporting of delegation and use [11,12]. These are complementary resources rather than interchangeable measures. The relevant evaluative question is whether documenting a contribution or verification step also demonstrates the capacity to justify a particular inference. Table 3 distinguishes the requirements of each instrument from this paper’s interpretation of the remaining assessment question.

Table 3. Governance instruments and the proposed assessment question [3,4,8-12].

Governance instru- ment	What it records or requires	Assessment question proposed in this viewpoint ^a
ICMJE ^b authorship criteria [8]	Substantial contribution; drafting or critical review; final ap- proval; agreement to be accountable for all aspects of the work.	Approval and agreement to be accountable can be given for system-generated content without comprehension or evaluation; accountability is an undertaking, not a tested capacity.
ICMJE guidance on AI use and COPE ^c position statement [3,4]	AI cannot be an author; use must be disclosed; humans remain responsible and should review AI-generated content.	Records that a tool was used and assigns responsibility, but does not assess the competence needed to discharge it.
CRediT ^d taxonomy and MeRiT ^e [9,10]	Named contributors mapped to discrete research roles or methodological steps.	Activity and attribution are reported; the depth of competence needed to justify a consequential contribution is not directly assessed by the role label.
GAIDeT ^f [11]	Delegated tasks classified across research activities; the stan- dard declaration states that tasks were delegated “under full human supervision” and that responsibility lies with the au- thors.	Supervision and responsibility are attested, not assessed; a task can be accurately classified without indicating whether the au- thors can justify its output.
GAMER ^g [12]	Nine reporting items, including the tool’s role in the study, AI-assisted manuscript sections, content verification, and im- pact on conclusions.	Verification and effects on conclusions are reported as activities and effects; the authors’ capacity to justify AI-derived content without relying on generative AI is not tested.

^aThe final column is the author’s analysis, not wording taken from the instruments.

^bICMJE: International Committee of Medical Journal Editors.

^cCOPE: Committee on Publication Ethics.

^dCRediT: Contributor Roles Taxonomy.

^eMeRiT: Method Reporting with Initials for Transparency.

^fGAIDeT: Generative Artificial Intelligence Delegation Taxonomy.

^gGAMER: Generative Artificial Intelligence Tools in Medical Research.

The ICMJE defines authorship through contribution, drafting or critical revision, final approval, and accountability for addressing questions about accuracy or integrity [8]. Its AI guidance requires disclosure, human responsibility, and review of generated content [3]. COPE likewise excludes AI tools from authorship and calls for transparent reporting [4]. These expectations already impose substantive duties; they should not be represented as merely administrative declarations. The narrower observation made here is that a reporting requirement does not itself supply a direct assessment of the competence

through which those duties are discharged. That distinction concerns implementation, not an alleged absence of responsibility norms.

CRediT and MeRiT identify who contributed and where, but their basic units are roles and activities [9,10]. This is valuable for tracing contributions and may help reveal the relevant expertise. It does not, by itself, establish the depth of understanding behind an adopted interpretation. Similarly, manuscript-level disclosure can make AI use visible without

resolving the content-specific question. This is a limitation of what can be inferred from the report, not evidence that the instruments endorse superficial evaluation. A competence assessment should therefore build on existing reporting rather than replacing it or treating a declaration as evidence of failure.

Recent AI frameworks move beyond a simple yes/no declaration. GAIDeT classifies delegated research tasks and connects their reporting with human supervision [11]. GAMER contains nine reporting items, including verification of content and effects on conclusions [12]. These requirements can support scrutiny of the evidential basis of a manuscript. The additional proposal here is to identify which author or authors can justify each consequential AI-assisted operation when a specific concern arises. Reporting and competence may coincide in practice; the point is that one should not infer the latter solely from the existence of the former.

Toward an Operational Governance Framework

The following instruments are proposals for extending contribution and AI-use reporting [10-12]. Their purpose is to make retained justificatory competence examinable, while limiting unnecessary administrative burden. The design is risk-graduated because a demand to document every iterative

interaction would be difficult to justify for purely presentational assistance. No evidence is offered that the proposed balance has already been achieved. The profiles, categories, decision rules, and safeguards should be treated as normative prototypes to be revised through pilots rather than as validated scales or existing journal requirements.

Epistemic Participation Profile, Extending CRediT, MeRIT, and GAIDeT

The proposed epistemic participation profile adds a competence description to a delegated task recorded through GAIDeT while identifying responsible authors in the spirit of MeRIT and CRediT [9-11]. The unit is a research operation in which AI supplied substantive inferential content rather than the manuscript as a whole. Table 4 specifies risks, responsible authors, evidence of evaluation, and a provisional rating; the particular risk categories in Table 4 remain proposals for piloting. Ratings are not aggregated: competence on peripheral operations cannot compensate for an uncovered central inference. The profile is a structured self-report subject to proportionate editorial scrutiny, not an objective measurement produced by the numerical labels. Table 5 provides a hypothetical completed example and a variant that fails the proposed threshold.

Table 4. Prototype epistemic participation profile extending task and contribution reporting (risk categories, ratings, and decision rules are proposed and unvalidated) [9-11].

Element	Specification
Unit of assessment	Each research operation in which AI supplied substantial inferential content, classified by GAIDeT ^a task category.
Risk level	Defined by the centrality of the dependent claims and the consequences of error: - Low: peripheral inferential content whose failure would not materially alter the manuscript's conclusions. - Moderate: content supporting secondary claims or methodological choices whose failure would weaken but not overturn the main conclusions. - High: content on which a primary conclusion rests, or with direct clinical, policy, or safety implications. Presentational assistance falls outside the profile and requires only manuscript-level disclosure.
Detectability	Whether an error is likely to be detected is recorded for interpreting the case; this provisional consideration does not reduce the claim's consequence-based risk category.
Responsible authors	Initials of the author or authors who meet the peer-challenge standard for the operation (MeRIT ^b style).
Rating 2: integrated	Named authors can justify adoption through evidence they can assess, including independent reanalysis or appropriately established reliability, and can answer reasonable specialist challenges. Another generated answer alone is not independent verification.
Rating 1: partially integrated	The core capacities are adequate for a limited low- or moderate-risk use but do not justify a stronger or higher-risk claim. Authors state the scope and limits of verification. Sampling alone does not determine the rating; its adequacy must be justified.
Rating 0: substituted	The authors can accept or reject the output but cannot explain its grounds or evaluate it independently.
Evidence	Verification performed (eg, reanalysis in independent software, checking of all cited sources, specialist consultation) and its scope.
Uncertainty	If the authors are uncertain between two levels, the lower level is recorded.
Required rating	High risk: 2; moderate risk: 1 or 2, with verification described; low risk: 1 or 2.
Below the required rating	Disclosure does not cure the deficit. The operation must be independently re-evaluated until the required rating is reached, re-assigned to an author with the relevant competence who contributes substantively, narrowed, or excluded from the manuscript's claims.
Link to reporting and audit	Every listed operation is reported at the operation level; high-risk operations are also reported at the claim level, with the structured explanation described in the auditability section. A specific unresolved concern may then trigger a reasoning audit.
Assessor	Self-assessment is attested by the named authors and compiled by the corresponding author; the editor checks that each operation meets its required rating and examines the profile in detail only for claim-level reports or audits.

^aGAIDeT: Generative Artificial Intelligence Delegation Taxonomy.

^bMeRIT: Method Reporting with Initials for Transparency.

Table 5. Hypothetical completed participation profile (initials and operations are illustrative; ratings apply the proposed rules in Table 4).

Operation (GAIDeT ^a category)	Risk	Named author(s)	Rating	Evidence	Consequence
Writing and editing: The LLM ^b drafted a background paragraph summarizing established epidemiology of the condition.	Low	AB	2	AB checked each statement against the cited primary sources and corrected one mis-attributed figure.	Operation-level profile entry; no further reporting.
Data management (analysis): The LLM proposed the multilevel model and drafted the interpretation of the primary outcome.	High	CD	2	CD respecified and reran the model in independent software, checked its assumptions, and revised the interpretation.	Claim-level report; threshold met; no audit unless a reviewer raises a specific concern.
Same operation, variant without a statistician.	High	None	0	Output accepted after a plausibility check.	Threshold not met; disclosure does not cure the deficit; authors must restore competence (eg, a statistician coauthor who reanalyzes the data), narrow the claim, or withdraw it.

^aGAIDeT: Generative Artificial Intelligence Delegation Taxonomy.

^bLLM: large language model.

Tiered Cognitive Provenance Reporting

The proposed reporting model has three tiers, extending the delegation and verification information requested by GAIDeT and GAMER [11,12]. Manuscript-level disclosure is appropriate for presentational assistance that does not change substantive claims. Operation-level reporting applies when AI supplies hypotheses, syntheses, methodological choices, interpretations, or arguments. Claim-level documentation is reserved for high-risk operations, as defined in Table 4, and includes the justification for the dependent conclusion. These tiers specify an intended allocation of scrutiny rather than an empirically optimized workload. Pilots must establish whether the categories can be applied consistently and whether their informational benefits justify their burden.

A Standard of Reasoning Auditability

Reasoning auditability operationalizes the proposed peer-challenge standard. Accounts of scientific understanding emphasize more than the possession of a fluent explanation and allow understanding to vary in depth and context [16,17]. Drawing on that background, the proposed report asks authors to state the grounds of the conclusion and its relevant dependencies, assumptions, limitations, alternatives, and revision conditions. It also asks them to distinguish adopted suggestions from the evaluation that warranted adoption. This list is a normative specification, not a validated diagnostic checklist. The response may use documentary and computational support; the relevant competence is the ability to assess and own the

justification, not memory of an original prompt or generation sequence.

The escalation procedure is also proposed here, building on the verification emphasis of GAMER [12]. Routine reporting is the first step; high-risk operations additionally receive a structured claim-level explanation. Only a specific unresolved concern should lead the handling editor to request a targeted reasoning audit. The author team may respond collectively, provide sources or code, qualify the claim, or demonstrate grounds for reliance on a validated method. An audit should not presume misconduct or require an oral examination. Its purpose is to resolve a concrete justificatory concern, and an editor should record what evidence resolves or sustains that concern.

Safeguards and Due Process for Reasoning Audits

Unequal assessment is a material concern. Liang et al [29] found that AI-text detectors could misclassify writing by nonnative English speakers, and Amano et al [30] documented burdens associated with working in English. These findings do not establish bias in reasoning audits, which have not been validated. They nevertheless motivate safeguards against using language fluency as a proxy for competence. Table 6 therefore proposes written triggers, explicit criteria, language assistance, accommodations, confidentiality, a response opportunity, and appeal. Potential effects on neurodivergent researchers, junior researchers, and institutions with fewer resources are concerns to investigate, not demonstrated outcomes of this proposed procedure.

Table 6. Proposed safeguards for targeted reasoning audits, informed by concerns about language disadvantage and detector bias (effectiveness evaluation required) [29,30].

Safeguard	Specification
Trigger	Only the handling editor may initiate an audit, on a written concern from a reviewer or editor tied to a specific inference, analysis, or interpretation. AI text detection scores, writing style, or disclosure of AI use are not triggers on their own.
Evidence requested	A written explanation of the flagged inference addressing the elements of the auditability standard; authors may add code, data, notes, or version histories. Oral examination is not required.
Who responds	Any named author who holds competence for the operation under the participation profile; a team response is acceptable.
Assessment criteria	The same published criteria for all manuscripts, applied to substance rather than fluency, recall of prompts, or reproduction of AI output. Authors may use language assistance in preparing the response.
Accommodations	Reasonable extensions and alternative formats on request, including for disability or neurodivergence.
Confidentiality and retention	Seen only by the handling editor, the reviewer who raised the concern, and, if needed, one independent methods specialist; retained with the confidential manuscript record under the journal’s standard retention policy; not shared with institutions except through established misconduct procedures.
Outcomes	Concern resolved; claim revised, qualified, or removed; independent specialist review or additional coauthor expertise required; or rejection. An audit alone does not constitute a finding of misconduct.
Response and appeal	Authors see the specific concern and may respond before any decision and may appeal through the journal’s standard appeals process.

Remediation and Prevention

A finding of inadequate coverage should lead to proportionate action rather than disclosure alone. Building on the duties of verification and correction discussed in recent authorship ethics [18], four responses are proposed. First, authors may reconstruct

and independently evaluate the relevant grounds. Second, they may obtain substantive specialist contributions and, where all authorship criteria are met, add an appropriately qualified coauthor [8]. Third, they may narrow or qualify the claim to match the evidence and competence available. Fourth, they may decline to adopt or withdraw the unsupported inference. These

responses are not a rigid sequence: a high-risk unsupported claim should not remain in a manuscript while remediation is merely promised.

Substitution is reversible in this framework because it concerns the current justificatory relation to a claim, not its causal origin. An author who comes to understand, evaluate, and defend a generated interpretation may restore the required coverage. Conversely, a declaration of supervision cannot substitute for that work. Institutional responses should also consider incentives and access to expertise, given existing concerns about hypercompetition and publishing pressures [7,28]. Training, methodological support, and assessment practices attentive to substantive contribution are possible preventive measures. Their effectiveness for this particular problem remains to be tested.

Equity and Differentiated AI Use

Equity requires distinguishing the assistance provided from the competence retained. Amano et al' [30] survey documents additional burdens faced by researchers working in a nonnative language. Accordingly, the proposed framework distinguishes assistance that expresses reasoning already understood, assistance that introduces substantive reasoning requiring evaluation, and assistance used where specialist support is unavailable. These categories describe different contexts, not fixed judgments about users. Translation and editing can preserve full answerability, whereas an unnoticed substantive alteration can introduce a new obligation to evaluate. Scarcity of specialist support calls for institutional assistance and proportionality but cannot, by itself, justify a claim that no member of the team can adequately assess.

Implementation and Empirical Validation

The proposed instruments require empirical validation before routine use. Their relation to existing task and verification reporting provides a practical starting point [10-12] but does not validate their thresholds. Initial pilots should measure author and editor time, agreement between independent assessors, unresolved or unsupported audit concerns, reporting

completeness, and successful remediation. Evaluation should also examine differential burdens across language backgrounds, career stages, regions, disability, and disciplines, informed by evidence of existing language inequalities [30]. Assessors should be blinded to irrelevant identity cues, where feasible, and should use an independent reference assessment of the substantive claim. An audit that ultimately resolves a concern is not necessarily a false positive; pilot reports should distinguish justified inquiry from an erroneous finding of deficient competence.

Conclusion

Scientific responsibility does not require every author to reproduce all the reasoning distributed across a research community. It requires, on the account defended here, sufficient team-level capacity to justify adopted claims, including justified reliance on collaborators and validated tools [8,14]. Derivative epistemic agency describes movement toward substitution; responsibility asymmetry specifies the mismatch that arises when formal responsibility persists without adequate coverage of comprehension, evaluation, or substantive answerability. The contribution is an operational specification within existing concerns about scientific understanding and epistemic duty [5,18], rather than a claim that AI introduces dependence into an otherwise self-sufficient science. Its threshold is a defended normative proposal, not an empirically validated diagnostic standard.

The proposed participation profile, tiered reporting, targeted reasoning audits, and remediation procedures extend existing contribution and AI-use frameworks [10-12]. Their central practical question is who can justify a consequential inference and on what grounds. They should protect legitimate collaboration and language assistance, while making unsupported adoption easier to identify and correct. Feasibility, assessment reliability, benefit, and unequal burdens remain open empirical questions for limited pilots. The aim is to preserve the conditions under which scientific claims remain answerable to evidence and criticism through identifiable human authors.

Acknowledgments

The author thanks the editor and reviewers for constructive comments.

Funding

No financial support was received for this work.

Data Availability

Data sharing is not applicable to this paper as no datasets were generated or analyzed during this study.

Generative AI Use

The author declares that generative AI (ChatGPT-6 Astra medium and Claude Opus 5.5 medium) was used for validation; editing, proofreading, or language polishing; summarizing text; and formatting references according to GAIDeT (Generative Artificial Intelligence Delegation Taxonomy). The author has verified all outputs of the tools used in the preparation of this manuscript. Responsibility for the final manuscript lies entirely with the author. Generative AI tools are not listed as authors and do not bear responsibility for the final outcomes.

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. Liang W, Zhang Y, Wu Z, Lepp H, Ji W, Zhao X, et al. Quantifying large language model usage in scientific papers. *Nat Hum Behav.* Dec 04, 2025;9(12):2599-2609. [doi: [10.1038/s41562-025-02273-8](https://doi.org/10.1038/s41562-025-02273-8)] [Medline: [40760036](https://pubmed.ncbi.nlm.nih.gov/40760036/)]
3. International Committee of Medical Journal Editors. Use of AI by authors. Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals. International Committee of Medical Journal Editors. URL: <https://www.icmje.org/recommendations/browse/artificial-intelligence/ai-use-by-authors.html> [accessed 2026-10-19]
4. Authorship and AI tools. COPE position statement. Committee on Publication Ethics. URL: <https://publicationethics.org/guidance/cope-position/authorship-and-ai-tools> [accessed 2026-10-19]
5. Messeri L, Crockett MJ. Artificial intelligence and illusions of understanding in scientific research. *Nature.* Mar 06, 2024;627(8002):49-58. [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/)]
6. Santoni de Sio F, van den Hoven J. Meaningful human control over autonomous systems: a philosophical account. *Front Robot AI.* Feb 28, 2018;5:15. [FREE Full text] [doi: [10.3389/frobt.2018.00015](https://doi.org/10.3389/frobt.2018.00015)] [Medline: [33500902](https://pubmed.ncbi.nlm.nih.gov/33500902/)]
7. Edwards MA, Roy S. Academic research in the 21st century: maintaining scientific integrity in a climate of perverse incentives and hypercompetition. *Environ Eng Sci.* Jan 01, 2017;34(1):51-61. [FREE Full text] [doi: [10.1089/ees.2016.0223](https://doi.org/10.1089/ees.2016.0223)] [Medline: [28115824](https://pubmed.ncbi.nlm.nih.gov/28115824/)]
8. International Committee of Medical Journal Editors. Defining the role of authors and contributors. Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals. International Committee of Medical Journal Editors. URL: <https://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html> [accessed 2026-10-19]
9. National Information Standards Organization. CRediT contributor role taxonomy. National Information Standards Organization. URL: <https://credit.niso.org/> [accessed 2026-10-19]
10. Nakagawa S, Ivimey-Cook ER, Grainger MJ, O'Dea RE, Burke S, Drobnik SM, et al. Method Reporting with Initials for Transparency (MeRIT) promotes more granularity and accountability for author contributions. *Nat Commun.* Apr 03, 2023;14(1):1788. [doi: [10.1038/s41467-023-37039-1](https://doi.org/10.1038/s41467-023-37039-1)] [Medline: [37012240](https://pubmed.ncbi.nlm.nih.gov/37012240/)]
11. 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 08, 2026;33(3):2544331-2544327. [doi: [10.1080/08989621.2025.2544331](https://doi.org/10.1080/08989621.2025.2544331)] [Medline: [40781729](https://pubmed.ncbi.nlm.nih.gov/40781729/)]
12. Luo X, Tham YC, Giuffrè M, Ranisch R, Daher M, Lam K, et al. GAMER Working Group. Reporting guideline for the use of generative artificial intelligence tools in MEDical research: the GAMER statement. *BMJ Evid Based Med.* Dec 01, 2025;30(6):390-400. [FREE Full text] [doi: [10.1136/bmjebm-2025-113825](https://doi.org/10.1136/bmjebm-2025-113825)] [Medline: [40360239](https://pubmed.ncbi.nlm.nih.gov/40360239/)]
13. Heersmink R. Distributed cognition and distributed morality: agency, artifacts and systems. *Sci Eng Ethics.* Apr 5, 2017;23(2):431-448. [doi: [10.1007/s11948-016-9802-1](https://doi.org/10.1007/s11948-016-9802-1)] [Medline: [27380187](https://pubmed.ncbi.nlm.nih.gov/27380187/)]
14. Ortmann J. Of opaque oracles: epistemic dependence on AI in science poses no novel problems for social epistemology. *Synthese.* Feb 05, 2025;205(2):80. [doi: [10.1007/s11229-025-04930-x](https://doi.org/10.1007/s11229-025-04930-x)]
15. Bietti LM, Bangerter A. Will the widespread use of large language models in scientific writing undermine scientists' critical thinking? *PLoS Biol.* Jun 3, 2026;24(6):e3003801. [FREE Full text] [doi: [10.1371/journal.pbio.3003801](https://doi.org/10.1371/journal.pbio.3003801)] [Medline: [42234608](https://pubmed.ncbi.nlm.nih.gov/42234608/)]
16. de Regt HW. *Understanding Scientific Understanding*. New York, NY. Oxford University Press; 2017.
17. Khalifa K. *Understanding, Explanation, and Scientific Knowledge*. Cambridge, UK. Cambridge University Press; 2017.
18. Cassimiro Nascimento D, da Silva D. Identical errors, different responsibilities: epistemic duty in AI-assisted scientific writing. *AI Ethics.* Aug 26, 2026;6(5):484. [doi: [10.1007/s43681-026-01330-w](https://doi.org/10.1007/s43681-026-01330-w)]
19. Birhane A, Kasirzadeh A, Leslie D, Wachter S. Science in the age of large language models. *Nat Rev Phys.* Apr 26, 2023;5(5):277-280. [doi: [10.1038/s42254-023-00581-4](https://doi.org/10.1038/s42254-023-00581-4)]
20. Lyell D, Magrabi F, Raban MZ, Pont L, Baysari MT, Day RO, et al. Automation bias in electronic prescribing. *BMC Med Inform Decis Mak.* Mar 16, 2017;17(1):28. [FREE Full text] [doi: [10.1186/s12911-017-0425-5](https://doi.org/10.1186/s12911-017-0425-5)] [Medline: [28302112](https://pubmed.ncbi.nlm.nih.gov/28302112/)]
21. Amrhein V, Greenland S, McShane B. Scientists rise up against statistical significance. *Nature.* Mar 20, 2019;567(7748):305-307. [doi: [10.1038/d41586-019-00857-9](https://doi.org/10.1038/d41586-019-00857-9)] [Medline: [30894741](https://pubmed.ncbi.nlm.nih.gov/30894741/)]
22. McNutt MK, Bradford M, Drazen JM, Hanson B, Howard B, Jamieson KH, et al. Transparency in authors' contributions and responsibilities to promote integrity in scientific publication. *Proc Natl Acad Sci U S A.* Mar 13, 2018;115(11):2557-2560. [FREE Full text] [doi: [10.1073/pnas.1715374115](https://doi.org/10.1073/pnas.1715374115)] [Medline: [29487213](https://pubmed.ncbi.nlm.nih.gov/29487213/)]
23. Gilbert SJ, Boldt A, Sachdeva C, Scarampi C, Tsai P. Outsourcing memory to external tools: a review of intention offloading. *Psychon Bull Rev.* Feb 05, 2023;30(1):60-76. [FREE Full text] [doi: [10.3758/s13423-022-02139-4](https://doi.org/10.3758/s13423-022-02139-4)] [Medline: [35789477](https://pubmed.ncbi.nlm.nih.gov/35789477/)]

24. Budzyń K, Romańczyk M, Kitala D, Kołodziej P, Bugajski M, Adami HO, et al. Endoscopist deskilling risk after exposure to artificial intelligence in colonoscopy: a multicentre, observational study. *Lancet Gastroenterol Hepatol* 2025;10(10):896-903. Correction to *Lancet Gastroenterol Hepatol* 2025; 10: 896-903. *Lancet Gastroenterol Hepatol*. Nov 2025;10(11):e12-903. [doi: [10.1016/S2468-1253\(25\)00294-8](https://doi.org/10.1016/S2468-1253(25)00294-8)] [Medline: [40946709](https://pubmed.ncbi.nlm.nih.gov/40946709/)]
25. Lee H, Sarkar A, Tankelevitch L, Drosos I, Rintel S, Banks R. The impact of generative AI on critical thinking: self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. 2025. Presented at: 2025 CHI Conference on Human Factors in Computing Systems; April 26-May 1, 2025:1-22; Yokohama, Japan. [doi: [10.1145/3706598.3713778](https://doi.org/10.1145/3706598.3713778)]
26. Elish MC. Moral crumple zones: cautionary tales in human-robot interaction. *Engag Sci Technol Soc*. Mar 23, 2019;5:40-60. [doi: [10.17351/ests2019.260](https://doi.org/10.17351/ests2019.260)]
27. Rubel A, Castro C, Pham A. Agency laundering and information technologies. *Ethic Theory Moral Prac*. Oct 24, 2019;22(4):1017-1041. [doi: [10.1007/s10677-019-10030-w](https://doi.org/10.1007/s10677-019-10030-w)]
28. Hanson M, Barreiro P, Crosetto P, Brockington D. The strain on scientific publishing. *Quant Sci Stud*. 2024;5(4):823-843. [doi: [10.1162/qss_a_00327](https://doi.org/10.1162/qss_a_00327)]
29. 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/)]
30. 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/)]

Abbreviations

COPE: Committee on Publication Ethics

CRedit: Contributor Roles Taxonomy

GAIDeT: Generative Artificial Intelligence Delegation Taxonomy

GAMER: Generative Artificial Intelligence Tools in Medical Research

ICMJE: International Committee of Medical Journal Editors

LLM: large language model

MeRIT: Method Reporting with Initials for Transparency

Edited by S Brini; submitted 30.Jun.2026; peer-reviewed by A Lunić, S Graham, O Ayilara, A Bobkov; comments to author 30.Sep.2026; revised version received 19.Oct.2026; accepted 21.Oct.2026; published 27.Oct.2026

Please cite as:

Elden Å

Derivative Epistemic Agency: AI-Mediated Scientific Cognition and the Limits of Human Answerability

J Meta and Res Int 2026;1:e105907

URL: <https://metascience.jmir.org/2026/1/e105907>

doi: [10.2196/105907](https://doi.org/10.2196/105907)

PMID:

©Åke Elden. Originally published in the JMIR Metascience and Research Integrity (<https://metascience.jmir.org>), 27.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 the 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.