The Human-in-the-Loop-O-Meter (HILOM)

A Trust-Based Framework for Disclosing AI Involvement in Scientific Research Publishing

Kevin A. Matthews, Ph.D.
Matthews Geographics LLC

April 2026

Acknowledgements

> *Developed using Claude Opus 4.6 (Anthropic), accessed via claude.ai, April 5, 2026, as drafting and synthesis tool; all core ideas, theoretical arguments, and source materials originated with the author. AI generated draft prose under detailed structural direction; author directed multiple restructurings, verified all content for accuracy, and the author's understanding of the framework evolved through the collaboration.*

Matthews, K.A. (2026). The Human-in-the-Loop-O-Meter (HILOM): A Trust-Based Framework for Disclosing AI Involvement in Scientific Research Publishing. Matthews Geographics LLC. matthewsgeographics.com

Abstract

Scientific publishers have responded to generative AI with a fragmented landscape of disclosure policies. All major publishers agree AI cannot be an author and humans bear full responsibility, but specific requirements diverge sharply on what triggers disclosure, where it must appear, how much detail is needed, and what AI use is permitted. This paper introduces the Human-in-the-Loop-O-Meter (HILOM), a structured, trust-based framework in which authors evaluate their work across eight dimensions covering the full research communication lifecycle. Each dimension corresponds to a specific risk created by how large language models work: LLMs default to statistically dominant patterns (Idea Origination), cannot trace their claims to sources (Research Contribution, Data Curation), diminish the cognitive work of articulation (Content Generation, Human Editing), cannot distinguish correct from incorrect output (Human Oversight), change the interpreter through interaction (Change in Thinking), and require a meta-commitment to honesty (Transparency). The self-assessment generates a narrative disclosure statement as its sole published output. HILOM is situated within the WCRIF-led effort to develop a Global Reporting Standard for AI Disclosure in Research and is designed to operate within the role-based attribution logic established by the CRediT taxonomy.

Keywords: AI disclosure, generative AI, scientific publishing, human-AI collaboration, authorship, trust-based framework, transparency, large language models, research ethics

Introduction

This paper introduces the Human-in-the-Loop-O-Meter (HILOM), a structured, trust-based framework designed to replace vague disclosure mandates with an eight-dimension self-assessment covering the full research communication lifecycle — from ideation through writing and disclosure. HILOM replaces binary declarations with graduated scales, replaces mechanical logging with structured self-assessment, and replaces external detection with an author-initiated social contract.

Two existing frameworks define what responsible, standardized attribution looks like in scholarly publishing and establish the direction toward which AI disclosure policy should be moving. The Contributor Role Taxonomy (CRediT), formalized as ANSI/NISO standard Z39.104-2022, provides a controlled vocabulary of fourteen contributor roles — from Conceptualization and Formal Analysis to Writing-Original Draft and Writing-Review and Editing — adopted by major publishers including Elsevier, Springer Nature, Wiley, and PLOS across thousands of journals [credit.niso.org]. CRediT's core insight is that attribution should be graduated, role-specific, and transparent rather than binary. It is the field's established model for human contributor disclosure, and its logic is the direct precedent for what AI disclosure should become. Alongside it, the World Conference on Research Integrity Foundation (WCRIF) is leading an active cross-organizational effort — with participation from the International Science Council, COPE, the STM Association, and the Global Young Academy — to develop a Global Reporting Standard for AI Disclosure in Research, with its first public consultation opened in February 2026. Together these initiatives define the destination: a publishing ecosystem in which every contribution, human and AI, is attributed consistently, structurally, and in a way readers can trust. HILOM is designed to operate within that trajectory. It does not compete with CRediT or the WCRIF standard — it fills the interval between where the industry is now and where those standards will take it.

Scientific publishers have responded to generative AI with a fragmented landscape of disclosure policies ranging from near-total bans to vague requests for acknowledgment. All major publishers and standards bodies agree on two narrow points: AI cannot be listed as an author, and human authors bear full responsibility for all content. Beyond this consensus, specific rules diverge sharply along four axes — disclosure trigger, disclosure location, detail level, and permitted scope — such that identical work faces different obligations depending on the target journal. The result is a governance vacuum that confuses authors, undermines transparency, and drives AI use underground. The current policy landscape, including disclosure requirements, AI image policies, and reviewer restrictions across major publishers, journals, conferences, standards bodies, and funding agencies, is documented and updated continuously at matthewsgeographics.com/policy-monitor.html.

Three structural failures explain why existing approaches fall short. First, binary disclosure conflates fundamentally different activities from grammar checking to hypothesis generation, producing statements that communicate nothing useful. Second, mechanical logging — requiring prompts, transcripts, or step-by-step records — assumes AI-assisted research follows a linear protocol. In practice, it is iterative and conversational, generating an infinite regress: each attempt to document AI use itself involves AI, requiring further documentation. The interpretive circle compounds this: interpretation permanently changes the interpreter, so the "original prompt" is a snapshot of a cognitive state that no longer exists. Third, detection tools produce false positives at rates unsuitable for high-stakes decisions and disproportionately flag non-native English speakers.

If external detection is unreliable and mechanical logging is structurally impossible, what remains is self-report. This is not a compromise but a recognition that authorship has always been trust-based: institutions trust first authors did the work, co-authors trust each other's contributions, and readers trust that data are real. A disclosure framework must therefore be built on the author's willingness to report honestly, and on the social contract between author and reader that the work was produced in good faith. This paper reports a conceptual framework-development process, not an empirical validation study.

Much of the discomfort researchers feel about AI disclosure stems not only from fragmented policies but from a lack of functional understanding of how large language models actually work. When researchers do not understand that LLMs generate text by predicting statistically probable sequences rather than reasoning or comprehending, disclosure requirements feel arbitrary rather than principled. Why does it matter who drafted the prose? Why does it matter whether the author verified the claims? These questions only have satisfying answers if the researcher understands the structural properties of the tool. HILOM addresses this by grounding each of its eight dimensions in a specific property of LLM architecture, making the why of disclosure as clear as the what.

The Proposed Framework

HILOM's primary output is a narrative disclosure statement generated through a self-assessment in which the author evaluates their work across eight dimensions. The assessment requires the author to first identify the AI tool by make, model, version, and access date. The author then evaluates their work on each dimension by selecting the descriptor that best characterizes their practice. These selections inform a narrative disclosure statement following a four-part structure: (1) what AI tool was used, (2) what the human originated, (3) what the AI contributed, and (4) how the human verified and transformed the output. HILOM does not prescribe what level of AI use is acceptable — that judgment belongs to the author, informed by disciplinary norms, publisher requirements, and their own ethical standards.

Table 1 summarizes all eight dimensions with their endpoint anchors. Full definitions, including all intermediate descriptors and examples, are provided in Appendix B.

Table 1. HILOM Dimension Summary with Endpoint Anchors.

#	Dimension	What It Measures	Fully Human Anchor	Maximum AI / Intensity Anchor
D1	Idea Origination	Who generated the core concepts, hypotheses, or creative direction	Human-originated: all direction from the author	AI-originated: author provided only broad topic
D2	Research Contribution	Who found, gathered, and synthesized the source material	Primarily human-led: traditional methods, minimal AI	AI-autonomous: AI conducted research with minimal direction
D3	Data Curation	Who prepared, cleaned, and validated empirical data. (N/A if non-empirical.)	Entirely human-led: all preparation done by human	AI-autonomous: AI prepared data with minimal involvement
D4	Content Generation	Who drafted the initial prose before editing	Entirely human-drafted: author composed all prose	Entirely AI-drafted: AI produced all prose from prompts
D5	Human Editing	How much the human transformed AI-drafted prose. (N/A if D4 = fully human.)	Minimal: grammar and punctuation fixes only	Transformative: AI output unrecognizable in final text
D6	Human Oversight	How deeply the human verified, validated, and scrutinized content	Surface-level: checked spelling and formatting only	Transformative authority: oversight reshaped output entirely
D7	Change in Thinking	How much the author's thinking changed through AI interaction	None: AI did not influence thinking	Fundamental: interaction reshaped understanding at a basic level
D8	Transparency	How openly and specifically the author discloses AI's role	Explicit and detailed: model named, process described	No disclosure: AI involved but not mentioned

The following is the HILOM disclosure statement for this paper (the complete dimension-by-dimension self-assessment is provided in Appendix A):

> *Developed using Claude Opus 4.6 (Anthropic), accessed via claude.ai, April 5, 2026, as drafting and synthesis tool; all core ideas, theoretical arguments, and source materials originated with the author. AI generated draft prose under detailed structural direction; author directed multiple restructurings, verified all content for accuracy, and the author's understanding of the framework evolved through the collaboration.*

The narrative disclosure statement can be produced in several ways. An author may write it directly from their completed self-assessment. Alternatively, an author may provide the completed assessment to an LLM and ask it to draft the narrative, or a future online tool could walk authors through the process and generate the narrative on their behalf. Regardless of how the narrative is produced, the human author must verify the final statement — just as with the HILOM framework itself, the disclosure is only as trustworthy as the author's willingness to ensure it accurately represents their work.

The Eight Dimensions

Idea Origination measures who generated the core concepts, hypotheses, or creative direction. The author assesses how much original direction came from their own thinking versus the model's statistical tendencies.

Research Contribution measures who found, gathered, and synthesized the source material. The human remains the only party capable of tracing claims to their origins and ensuring the factual foundation is sound.

Data Curation and Preparation measures who prepared, cleaned, validated, and structured the empirical data for reporting — not the execution of analyses, which belongs in the Methods section. Mark N/A for non-empirical work. (Disclosure: this dimension was suggested by Claude during drafting and refined by the author.)

Content Generation measures who drafted the initial prose before editing.

Human Editing and Refinement measures how much the human transformed AI-drafted prose after generation. This is where slop is filtered out and authenticity arises.

Human Oversight and Quality Control measures whether the author verified the substance of the content — distinct from Human Editing (D5), which measures transformation of expression. Only human oversight can prevent epistemic alienation.

Change in Thinking measures the degree to which the author's own understanding changed through AI interaction. (Disclosure: this dimension was suggested by Claude Sonnet 3.5 during drafting and refined by the author.)

Transparency of AI Use measures how openly and specifically the author discloses AI's role, including tool identification. HILOM advocates completing a disclosure statement even when no AI was used, because silence in an AI-saturated environment may be interpreted as evasion; an explicit non-use statement is itself a contribution to transparency.

Why These Eight Dimensions

A large language model generates text by predicting the most probable next word in a sequence, trained on enormous quantities of human-written text using transformer architecture. An LLM does not understand what it writes: an incorrect output carries exactly as much statistical confidence as a correct one. The distinction between information and knowledge is categorical, not a matter of degree. An LLM produces information — statistically plausible text. Converting that information into knowledge requires human contributions that are different in kind from what the model provides: asking the question that motivated the inquiry, evaluating whether claims are warranted, and determining whether the output represents what the author actually thinks. Each of HILOM's eight dimensions measures one of these contributions. Generative AI also means that content can now be produced faster than it can be read, shifting the bottleneck in scientific communication from production to evaluation. In this environment, disclosure of how humans evaluated and shaped AI-assisted work becomes more important than disclosure of whether AI was used at all.

The Novelty Problem

LLMs default to statistically dominant patterns in their training data. They reproduce what is common, not what is new. They also inherit the frequency biases of their training corpus: disciplinary terms with multiple meanings default to whichever usage dominates. For example, "rural health" in the biomedical literature predominantly refers to clinical healthcare delivery rather than the public health of rural populations, and an author relying on AI-generated text without recognizing this skew risks reproducing the dominant framing uncritically. Curiosity — the biologically grounded drive to resolve uncertainty — is what generates genuinely novel questions. Neuroscience identifies a primary exploratory drive in the brain's affective architecture that has no analog in language models: LLMs do not experience information gaps or feel uncertainty. The Idea Origination dimension tracks whether the driving question came from a curious human or from the model's statistical tendencies. The least defensible scenarios are those where the human outsources not just labor but the wanting-to-know.

The Provenance Problem

LLMs are provenance-blind: they cannot distinguish a synthesized insight from a near-paraphrase of a single source. The model does not know where its claims come from and cannot trace them back to primary evidence. This problem is compounded by a circular dynamic in scholarly publishing: major publishers — including Wiley, Taylor and Francis, and Oxford University Press — have signed multimillion-dollar licensing deals selling their published content to AI companies for model training. A researcher may therefore use an LLM trained on published papers, possibly including their own, and the LLM's output may carry traces of that content without attribution. When the researcher then publishes with AI assistance, readers have no way of knowing whether an AI-generated passage reflects genuine synthesis or near-reproduction of a single source. HILOM cannot solve this provenance problem — it is anchored in what the author did, not in the AI's internal processes — but the Research Contribution dimension requires the author to attest that they verified the factual and intellectual foundations of the work.

The Data Curation dimension extends this to the empirical pipeline: AI can clean, format, and restructure data at scale, but it does so without understanding what the data mean. A statistical outlier might be an error or the most important finding in the dataset. If AI silently drops records, recodes variables, or imputes missing values based on training-data patterns, those choices propagate through every downstream analysis without the author necessarily being aware. Errors at this stage are invisible in the final manuscript but can invalidate the entire study.

The Voice Problem

Articulating an idea in language is not merely recording a pre-existing thought; it is constituting the thought. When an LLM drafts prose, the word choices, sentence structures, and argumentative moves are drawn from corpus statistics, not from the author's reasoning. The resulting text expresses a different thought than what the author would have written. McLuhan observed that any technology extending a human capacity simultaneously diminishes the original ability: the car extends locomotion but diminishes the habit of walking; the calculator extends computation but diminishes mental arithmetic. Generative AI extends the capacity to produce written text while diminishing the cognitive work of articulation — which is precisely where thinking happens. The Content Generation dimension measures the degree of this trade-off: how much of the articulation was outsourced to the model.

The Human Editing and Refinement dimension is where the author reverses the trade-off. In common usage, "slop" has become the term for AI-generated content produced without meaningful human intent, oversight, or intervention — text that is optimized for statistical plausibility rather than authentic communication. A piece of writing may be grammatically correct and logically structured, but if it lacks the author's authentic voice and genuine intent, it fails as communication. The Human Editing dimension is where authenticity arises. It is also a direct counter to the Content Generation dimension which discloses who drafted the prose while the Human Editing dimension discloses whether the author verified that the result represents their actual ideas.

The Verification Problem

An incorrect LLM output carries exactly as much statistical confidence as a correct one. The model has no mechanism for self-correction and no capacity to flag its own errors. Whereas the Human Editing dimension asks whether the content accurately represents the author's ideas, the Human Oversight and Quality Control dimension asks whether the content is valid. A researcher can completely rewrite AI prose for style while never checking whether the underlying claims are factual. This distinction — between transforming expression and verifying substance — is HILOM's most important analytical contribution. Accepting AI output without deep processing produces what Lin calls "epistemic alienation": the author endorses content they have not fully evaluated. The three-dimension pipeline of Content Generation → Human Editing → Human Oversight asks, in sequence: who wrote it, does it represent me, and is it factually correct.

The Transformation Problem

Interaction with AI changes the interpreter. The Change in Thinking dimension measures the degree to which the author's own understanding changed through AI interaction. Disclosing cognitive transformation is an act of intellectual honesty, not an admission of weakness. A person reading a book can be influenced by its ideas, just as a person reviewing AI-generated content can be influenced by that content; the cognitive mechanism is identical. We would never require a researcher to disclose that a particular book changed how they think about a problem — yet HILOM asks for this disclosure because transparency about how ideas developed strengthens the trust relationship between author and reader. When AI introduces an idea the author incorporates, the author has a due diligence obligation to determine its provenance: was it synthesized across the training corpus, or effectively lifted from a single identifiable source deserving citation? This connects directly to the Research Contribution dimension, where the author assesses how thoroughly they verified source material.

The Trust Problem

The first dimension (Idea Origination) through the Change in Thinking dimension measures what happened. The Transparency of AI Use dimension measures whether the author is willing to tell the reader what happened. It is the meta-dimension that makes the other seven meaningful: without it, the assessments are unverifiable. HILOM advocates completing a disclosure statement even when no AI was used, because in an environment where AI-generated content is common, silence may be interpreted as evasion. An explicit non-use statement is itself a contribution to transparency.

These six problems — novelty, provenance, voice, verification, transformation, and trust — are not independent. They interact: an author who outsources Idea Origination (D1) and never verifies the AI's sources (Research Contribution, D2; Human Oversight, D6) while failing to recognize how the interaction changed their thinking (Change in Thinking, D7) has produced work that may be statistically plausible but epistemically hollow. The due diligence required by HILOM — verifying sources, checking factual accuracy, determining the provenance of AI-suggested ideas, filtering out slop, reclaiming authorial voice — may not save the author time. The efficiency gains of AI drafting can be offset or exceeded by the verification burden of responsible use. HILOM makes this trade-off visible rather than leaving it hidden.

Discussion

HILOM provides the structural architecture that is missing from the current AI disclosure landscape: a standardized self-assessment producing comparable narrative statements across journals, disciplines, and levels of AI involvement. This version produces a narrative disclosure statement as its sole output, using descriptive anchors rather than coded labels for each dimension. This design prioritizes clarity and usability over machine-readability. However, if HILOM is adopted at scale — for example, in meta-research examining AI-use patterns across a corpus of published papers — a standardized, machine-readable coding scheme would enable quantitative aggregation, cross-study comparison, and longitudinal tracking. Future versions should therefore develop a companion coding system with consistent notation across all eight dimensions, designed for computational processing while mapping directly to the descriptive anchors used in the narrative assessment.

AI tools can meaningfully reduce barriers for non-native English speakers producing work for English-language journals, and generative AI functions as assistive technology for neurodivergent researchers — including those with ADHD, dyslexia, and other conditions affecting writing fluency, executive function, and text production — by supporting academic writing, task planning, and self-management. Policies that restrict or stigmatize AI use disproportionately affect these populations. Because these researchers benefit most from AI tools, they especially need a disclosure framework that centers what the human contributed rather than stigmatizing the tool. HILOM's trust-based design is intended to serve this function.

Several prior and concurrent efforts address AI disclosure or reporting in scholarly work. Table 2 compares the principal frameworks, identifying what each contributes and what HILOM adds. The coverage gap analysis reveals four dimensions no existing framework adequately captures: idea origination as a separately scored measure, the editing-versus-oversight distinction, change in thinking, and data curation in the context of manuscript production.

Table 2. Comparison of AI Disclosure and Reporting Frameworks.

Framework	Type	Scope	Key Feature	What HILOM Adds
AID [Weaver, 2024]	Role-based checklist	Writing/contribution roles (CRediT model)	Narrative AID Statement	Idea origination, change in thinking, scored oversight, proportional scales
Cognitio Emergens [Lin, 2025]	Theoretical	Human-AI knowledge co-creation	Agency Configurations; epistemic alienation concept	Practical disclosure output; structured self-assessment
CANGARU [Cacciamani et al., 2023]	Reporting standard	Cross-journal standardization	Addresses 'Babel Tower Effect'	Author-facing self-assessment generating required information
CONSORT-AI / SPIRIT-AI [Chen et al., 2025]	Clinical trial extensions	AI as clinical intervention	Adapts trial reporting standards	AI used in research production, not AI studied as intervention
GAMER [Luo et al., 2025]	Reporting checklist	GenAI in medical research	Nine-item checklist	Self-assessment generating answers, not post-hoc checklist
STM Classification [STM, 2025]	Activity taxonomy	AI in manuscript preparation	Nine categories for publishers	Scored degree of involvement, not categorical presence/absence
GAIDeT [Suchikova et al., 2025]	Delegation taxonomy	AI task delegation across research	Macro/micro checklist with generator	Editing/oversight distinction, change in thinking, narrative output
HILOM (this paper)	Self-assessment	Full research lifecycle	Eight scored dimensions + narrative	N/A

HILOM, CRediT, and the WCRIF Standard

HILOM's relationship to the WCRIF Global Reporting Standard and the CRediT taxonomy is worth stating plainly, because it clarifies both what HILOM is and what it is not trying to be. The most immediate relationship is temporal: HILOM is a stopgap. The WCRIF standard-setting process is underway, but finalizing a global standard, securing publisher adoption, and embedding it in submission systems across thousands of journals is a multi-year undertaking.

Authors navigating the fragmented policy landscape cannot wait. HILOM fills that interval, requiring nothing of publishers except acceptance of a narrative disclosure statement — which virtually all existing policies already accommodate.

But HILOM also offers something that neither CRediT nor the WCRIF process has yet resolved: a practical output. CRediT's fourteen roles, applied systematically to AI involvement across a research project, would in principle require documenting AI contributions to Conceptualization, Data Curation, Formal Analysis, Writing-Original Draft, Writing-Review and Editing, and any other applicable role — each with sufficient specificity to be meaningful. In practice, that is not a disclosure statement; it is a methods section. HILOM collapses the same intellectual territory into a single paragraph by assessing eight dimensions in sequence and generating one narrative output — concise enough to appear in a manuscript's Acknowledgments, specific enough to satisfy any current journal policy, and structured enough to feed directly into the machine-readable standards the WCRIF process will eventually require.

A third relationship is architectural: HILOM's dimension structure is compatible with the role-based logic CRediT has already established. Idea Origination is not Content Generation, which is not Human Oversight — for precisely the same reason that Conceptualization is not Writing-Original Draft. Any global standard will need to make those distinctions. HILOM demonstrates that authors can make them, reliably, in plain language, without compliance infrastructure. That is the evidence base the WCRIF process needs.

Limitations

HILOM is trust-based with no enforcement mechanism. However, authorship itself has never been externally enforced: no one verifies in real time that first authors did the work or that data are real. HILOM makes the trust that has always been implicit into something explicit and structured; enforcement-focused approaches drive AI use underground.

No formal user testing, inter-rater reliability assessment, or empirical validation was conducted. Priorities include inter-rater reliability testing, test-retest stability studies, editorial usability evaluations, cross-disciplinary testing, and calibration exercises generating benchmark scenarios with anchor examples and scoring guidance.

Several dimensions — particularly Idea Origination, Human Editing, and Change in Thinking — require subjective judgments that may vary across authors. The alternative is excluding these dimensions, producing consistency at the cost of capturing only the least informative aspects of AI use. Authors may forget steps, interpret dimensions differently, or underreport strategically. Optional corroborating artifacts — such as conversation logs, version histories, or timestamped drafts — can provide selective verification where feasible.

HILOM addresses the author side of the disclosure problem. It does not address a structural challenge that the publication industry must confront independently: AI has lowered the cost of manuscript production without reducing the human cost of review. The 2018 Publons Global State of Peer Review documented that editors already needed to send an average of 2.7 invitations per accepted review, with projections estimating 3.6 by 2025. More recently, 21% of peer reviews submitted to the International Conference on Learning Representations (ICLR) were found to have been AI-generated, and NeurIPS 2025 received more than 27,000 submissions — a volume that, at even 30 minutes per review, would require more than 13,500 person-hours of reviewer effort. Any disclosure policy that makes submission easier without addressing review capacity will intensify this imbalance. That problem is outside HILOM's scope, but it is not outside the industry's responsibility.

Conclusion

HILOM addresses a structural gap in the current landscape: publishers require disclosure but provide no standard structure for producing it, authors want to comply but face contradictory instructions across journals, and readers receive binary statements that communicate nothing useful about the nature or depth of AI involvement. By centering governance on what the human contributed — across eight dimensions covering ideation, research, data curation, drafting, editing, oversight, cognitive transformation, and transparency — HILOM provides the missing architecture for a trust-based disclosure that is both principled and practical.

The framework's implications extend beyond any individual disclosure statement. The fragmented policy landscape is a symptom of a system that attempted to govern a new technology through prohibition and surveillance before it understood what it was governing. Binary disclosure, prompt logs, and detection tools share a common assumption: that AI use is a problem to be caught rather than a practice to be understood. That assumption has produced policies that are structurally unenforceable, that stigmatize good-faith authors, and that generate compliance theater rather than transparency. The field's most productive path forward is not more surveillance but clearer expectations, standardized structure, and recognition that the social contract between author and reader — which has always been trust-based — is sufficient to carry the weight of AI disclosure when authors are given the tools to fulfill it.

HILOM offers three contributions to that path. As a stopgap, it gives authors a principled instrument for right now, before any global standard is finalized. As an evidence base, it demonstrates that graduated, role-specific AI disclosure is feasible, comprehensible, and compatible with the attribution logic CRediT has already established as the field's preferred model. And as a design proposal, it surfaces the specific distinctions — between origination and generation, between editing and oversight, between using AI and being changed by it — that any global standard will need to encode. The WCRIF harmonization process is the appropriate venue for that encoding. HILOM is designed to be useful to it.

References

BaHammam, A. S. (2025). The Transparency Paradox: Why Researchers Avoid Disclosing AI Assistance in Scientific Writing. *Nature and Science of Sleep, 17*, 2569-2574. https://doi.org/10.2147/NSS.S568375

Bender, E. M., Gebru, T., McMillan-Major, A., & Mitchell, M. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610-623. https://doi.org/10.1145/3442188.3445922

Cacciamani, G. E., Eppler, M., Ganjavi, C., & Pekan, A. (2023). Development of the CANGARU Guidelines. *arXiv*. https://doi.org/10.48550/arXiv.2307.08974

Chen, D., He, E., Pace, K., Chekay, M., & Raman, S. (2025a). Concordance with CONSORT-AI guidelines. *BMJ Oncology, 4*(1), e000733. https://doi.org/10.1136/bmjonc-2025-000733

Chen, D., He, E., Pace, K., Chekay, M., & Raman, S. (2025b). Concordance with SPIRIT-AI guidelines. *The Oncologist, 30*(5), oyaf112. https://doi.org/10.1093/oncolo/oyaf112

Clarivate. (2018). *Publons Global State of Peer Review 2018*. https://publons.com/static/Publons-Global-State-Of-Peer-Review-2018.pdf

COPE Council. (2023). COPE position: Authorship and AI tools. https://publicationethics.org/guidance/cope-position/authorship-and-ai-tools

da Veiga, A. (2025). Ethical guidelines for the use of generative AI in scholarly publishing. *Science Editing, 12*(1), 28-34. https://doi.org/10.6087/kcse.352

Flanagin, A., Bibbins-Domingo, K., Berkwits, M., & Christiansen, S. L. (2024). Reporting Use of AI in Research and Scholarly Publication — JAMA Network Guidance. *JAMA, 331*(13), 1096-1098. https://doi.org/10.1001/jama.2024.3471

George, E. O. (2025, November 7). Understanding AI Disclosures. Paperpal.
Heidegger, M. (1927). Being and Time. Trans. J. Macquarrie & E. Robinson. Harper & Row, 1962.

ICMJE. (2024, January). Recommendations for Scholarly Work in Medical Journals. https://www.icmje.org/recommendations/browse/artificial-intelligence/ai-use-by-authors.html

International Science Council. (2026). AI disclosure in research: Towards a global reporting standard.

Kwon, D. (2024, December 9). Publishers are selling papers to train AIs — and making millions of dollars. *Nature*. https://doi.org/10.1038/d41586-024-04018-5

Lin, X. (2025). Cognitio Emergens. https://arxiv.org/abs/2505.03105

Loewenstein, G. (1994). The psychology of curiosity. *Psychological Bulletin, 116*(1), 75-98. https://doi.org/10.1037/0033-2909.116.1.75

Luo, X., Tham, Y. C., Giuffre, M., et al. (2025). The GAMER Statement. *BMJ Evidence-Based Medicine*. https://doi.org/10.1136/bmjebm-2025-113825

Matthews, K. A., Spears, K. S., & Anderson-Lewis, C. (2025). Rural Health Disparities: Contemporary Solutions for Persistent Rural Public Health Challenges. *Preventing Chronic Disease, 22*, E27. https://doi.org/10.5888/pcd22.250202

McLuhan, M. (1964). Understanding Media: The Extensions of Man. McGraw-Hill.
National Information Standards Organization. (2022). *ANSI/NISO Z39.104-2022, CRediT, Contributor Roles Taxonomy*. https://credit.niso.org

Nature. (2025, December 1). Major AI conference flooded with peer reviews written fully by AI. https://www.nature.com/articles/d41586-025-03506-6

Panksepp, J. (1998). Affective Neuroscience. Oxford University Press.

PMC. (2024). The challenge of reviewers scarcity in academic journals: payment as a viable solution. PMC11461020.

So, R. (2025, November 8). Publisher Guidelines on AI-Assisted Scholarly Writing. Project Rachel.

STM Association. (2025, September). Recommendations for a Classification of AI Use in Academic Manuscript Preparation.

Suchikova, Y., Tsybuliak, N., Teixeira da Silva, J. A., & Nazarovets, S. (2025). GAIDeT. *Accountability in Research*. https://doi.org/10.1080/08989621.2025.2544331

Vaswani, A., et al. (2017). Attention Is All You Need. NeurIPS, 30. arXiv:1706.03762.

Weaver, K. D. (2024). The AID Framework. *C&RL News, 85*(10). arXiv:2408.01904.

Wittgenstein, L. (1953). Philosophical Investigations. Trans. G. E. M. Anscombe. Blackwell.
World Conference on Research Integrity Foundation (WCRIF). (2026). Focus track: Global Reporting Standard for AI Disclosure in Research.

Yoo, J.-H. (2025). Defining the Boundaries of AI Use in Scientific Writing. *JKMS, 40*(23), e187. https://doi.org/10.3346/jkms.2025.40.23.e187

Zhao, X., Cox, A., & Chen, X. (2025). The use of generative AI by students with disabilities in higher education. *The Internet and Higher Education, 66*, 101014.

Zielinski, C., et al. (2023). Chatbots, generative AI, and scholarly manuscripts: WAME recommendations. *Colombia Medica, 54*(3), e1015868.

Appendix A: Author Self-Assessment for This Paper

AI Tool: Claude 3.5 Sonnet (Anthropic), accessed via claude.ai, February 2026.

Dimension	Descriptor Selected	Explanation
Idea Origination	Human-driven	HILOM framework, philosophical grounding, and argument for trust-based disclosure all originated with the author. AI suggested D3 and D7 dimensions, which the author evaluated and adopted.
Research Contribution	Human-led with AI assistance	Author assembled 38-source corpus and directed all research. AI synthesized summaries and helped map sources to themes; author verified all sources.
Data Curation	Not applicable	Theoretical paper. No empirical data collected or analyzed.
Content Generation	Mostly AI-drafted	AI generated majority of draft prose from detailed outlines and structural direction provided by author.
Human Editing	Substantial	Author directed two complete restructurings and provided detailed feedback reshaping argument, organization, and emphasis.
Human Oversight	Ethical and contextual scrutiny	Author reviewed all content for accuracy, verified citations against source materials, and evaluated conceptual consistency with domain expertise.
Change in Thinking	Significant	Author's understanding of the framework evolved through collaboration.
Transparency	Explicit and detailed	Full disclosure with AI tool named, process described, and detailed breakdown provided.

Appendix B: Full Dimension Scale Definitions

This appendix provides the complete scale definitions for each of HILOM's eight dimensions. Table 1 in the body of the paper provides endpoint anchors; these tables provide all intermediate descriptors with examples. Authors should select the descriptor that best characterizes their practice on each dimension.

Table B.1. Idea Origination

Descriptor	Definition	Example
Human-originated	All research questions, hypotheses, or creative direction came from the author.	Scientist develops theory from own observations.
Human-driven	Author generated core ideas; AI suggested alternatives the author evaluated.	Writer has concept; asks AI for sub-plot ideas, selects one.
Collaborative	Both contributed substantially; author set parameters and made final decisions.	Team gives AI broad questions; AI suggests directions; team selects.
AI-augmented	AI identified patterns that became central; author validated and directed.	AI finds unexpected pattern; scientist designs experiments around it.
AI-originated	Author provided only broad topic; AI generated specific questions and hypotheses.	One-sentence prompt; AI generates project concepts.

Table B.2. Research Contribution

Descriptor	Definition	Example
Primarily human-led	Traditional methods; AI used minimally.	Library research; AI summarizes one article.
Human-led with AI assistance	Human chose topics/sources; AI helped find and summarize.	Researcher selects key authors; AI finds papers; human verifies.
Balanced	Human set questions; AI synthesized across sources with guidance.	Human poses questions; AI drafts lit review; human heavily revises.
Primarily AI-led	AI did most retrieval and synthesis; human verified.	AI creates knowledge summary; human fact-checks.
AI-autonomous	AI conducted research with minimal direction.	Human requests full report; AI searches, synthesizes, writes.

Table B.3. Data Curation (N/A for non-empirical work)

Descriptor	Definition	Example
Not applicable	Non-empirical work.	Theoretical paper or literature review.
Entirely human-led	All data cleaning, validation, structuring done by human.	Researcher manually cleans survey data.
Human-led with AI assistance	AI helped with routine tasks; human directed.	AI scripts standardize format; human validates.
Balanced	AI performed substantial processing under human oversight.	AI flags outliers; human reviews each flag.
Primarily AI-led	AI handled most preparation; human provided high-level review.	AI ingests and cleans large dataset; human spot-checks.
AI-autonomous	AI prepared data with minimal human involvement.	AI runs pipeline; human reviews only final output.

Table B.4. Content Generation

Descriptor	Definition	Example
Entirely human-drafted	Author composed all prose without AI.	Writer types full article from scratch.
Mostly human-drafted	AI contributed only isolated sentences.	Student writes essay; uses AI for alternate phrasings.
Jointly drafted	Both produced significant portions of first-draft prose.	Creator writes intro/conclusion; AI drafts middle sections.
Mostly AI-drafted	AI produced majority from author's outlines.	Detailed outline provided; AI generates full prose.
Entirely AI-drafted	AI produced all prose from human prompts.	Comprehensive prompt; AI produces complete draft.

Table B.5. Human Editing (N/A if Content Generation Dimension = ‘Entirely human-drafted’

Descriptor	Definition	Example
Not applicable	Human wrote all content.	Entirely human-authored work.
Minimal	Grammar, spelling, punctuation fixes only.	Fixes typos in AI-drafted email.
Moderate	Rephrased sentences, improved clarity and flow.	Rewrites sentences and combines paragraphs.
Substantial	Major rewriting, reorganization, new content added.	Rewrites dialogue, changes structure, adds details.
Transformative	AI output only a starting point; result unrecognizable.	Philosopher reconstructs argument entirely.

Table B.6. Human Oversight (D6).

Descriptor	Definition	Example
Surface-level review	Checked spelling, grammar, formatting only.	Quick read for typos.
Clarity and conciseness	Ensured text is readable and efficient.	Reviewed for accessibility.
Content and consistency	Fact-checked claims; verified logical consistency.	Checks every statistic against original data.
Ethical and contextual scrutiny	Applied expert judgment and domain knowledge.	Reviews syllabus for bias and teaching values.
Transformative authority	Oversight reshaped output; added original insights.	Critiques and reconstructs argument.

Table B.7. Change in Thinking

Descriptor	Definition	Example
None	AI did not influence thinking. Mechanical tasks only.	Grammar checking only.
Low	AI helped articulate existing ideas. Core thinking unchanged.	AI rephrased existing arguments more clearly.
Moderate	AI introduced connections or framings not previously considered.	AI suggested theoretical connection author incorporated.
Significant	Understanding changed substantially through AI interaction.	Through iterative dialogue, reframed research question.
Fundamental	Interaction reshaped understanding at a basic level.	Entered with one framework, emerged with another.

Table B.8. Transparency

Descriptor	Definition	Example
Not applicable	No AI was used. Completing this statement is recommended.	Entirely human-authored work with explicit non-use statement.
Explicit and detailed	Names model/version, describes process and extent.	Dedicated AI Assistance Disclosure section.
General disclosure	Mentions AI use with limited specifics.	'AI tools assisted in preparation.'
Implicit disclosure	Hints at AI involvement without stating it.	Thanks 'digital research tools' without detail.
No disclosure	AI was involved but not mentioned.	Uses AI extensively; submits as entirely human.

Appendix C: Policy Monitor Prompt

Run this prompt in your chosen LLM platform for the most current information about AI disclosure policies in peer-reviewed scientific publishing.

You specialize in academic publishing policy, research ethics, and research-funder governance, with deep expertise in AI/GenAI disclosure requirements across scholarly publishing, standards bodies, conferences, and funding agencies. You are helping me build institutional guidance for researchers at my organization by producing an up-to-date reference on AI disclosure policies.

The Human-in-the-Loop-O-Meter: AI Disclosure Policies in Peer Reviewed Scientific Publication is a framework published by Matthews Geographics, LLC (matthewsgeographics.com). That white paper contains a structured policy table covering academic, publishing, conference, standards, and funding entities' AI policies.

Task

Research and verify the latest AI disclosure, authorship, image-use, and peer-review/applicant restrictions for each entity listed below.

Before building the table: (1) Conduct a brief discovery scan to identify any additional major entities that now have explicit public AI/GenAI policies relevant to scholarly publication or grant review. (2) Add newly identified entities only if they have a formal public-facing policy, guidance page, or official instructions affecting authors, reviewers, editors, conference submitters, or grant applicants/reviewers. (3) Distinguish clearly between: formal policy or enforceable submission/review guidance; journal-family or publisher guidance; emerging standards initiatives or in-development frameworks.

Deliverables

Produce four outputs in this order:

A. Newly identified entities. A short bullet list of any major additional entities found beyond the seed list, with one-sentence justification for inclusion.

B. Updated six-column policy table. Produce an updated six-column table preserving these dimensions exactly: Scope, Trigger, Location, Detail, AI Images, Peer Reviewer AI restrictions.

C. Change log. Briefly note: entities added; entities removed or merged; rows materially changed from prior assumptions; rows where the policy is ambiguous or still evolving.

D. Excluded-but-noted entities. List any entities scanned but not included because they do not yet have a sufficiently explicit or stable AI policy.

Core Entities to Research

ICMJE, COPE, WAME, WCRIF/WCRI, Council of Science Editors (CSE), Springer Nature, Elsevier, Wiley, Taylor and Francis, SAGE Publishing, Oxford University Press, Cambridge University Press and Assessment, Cell Press, Frontiers, MDPI, Science (AAAS), Nature Portfolio, PLOS ONE, ACS, IEEE, ACM, APA, AGU, ASM Journals, PNAS, The Lancet, NEJM, BMJ, JAMA Network, NeurIPS, ICML, ICLR, NIH, NSF, UK Research and Innovation (UKRI), Wellcome.

Research Rules

Use official primary sources only when possible: publisher policy pages, author instructions, reviewer instructions, conference policy pages, funder notices, or standards-body guidance. Prefer policy-specific URLs over homepages. Verify whether the policy applies at the publisher level, journal-family level, specific-journal level, conference level, or funder/review-process level. Capture exact disclosure location when stated (e.g., Methods, Acknowledgments, cover letter, checklist, submission form, project description). Capture whether the trigger is broad disclosure, only substantive use, only generated text/images, only non-language-editing use, or another threshold. Capture whether AI-generated images are prohibited, permitted with disclosure, reviewed case-by-case, editor-approved only, or not addressed. Capture reviewer restrictions precisely, including whether AI use is prohibited, restricted due to confidentiality, allowed only in limited internal workflows, or not addressed. Note when a body is issuing principles or recommendations rather than binding policy. Note publication/update date when available. Flag policies that appear to be in development, fragmented across journals, or inconsistent across sub-brands.

Output Quality Requirements

Be precise and conservative. Do not overstate. Do not infer permissions that are not explicitly stated. When policies differ across sub-journals or sub-brands, say so. When no clear AI policy is found, say "No explicit public AI policy located" rather than guessing. Include a source URL for every row in a separate "Source" field after the table, or in a structured source list keyed to entities. End with a short synthesis identifying the biggest cross-entity patterns and the most important unresolved ambiguities.

Suggested Citation

Matthews, K. A. (2026). The Human-in-the-Loop-O-Meter (HILOM): A Trust-Based Framework for Disclosing AI Involvement in Scientific Research Publishing. Matthews Geographics LLC. matthewsgeographics.com

About Matthews Geographics

Matthews Geographics LLC is an independent consulting practice specializing in democracy infrastructure, civic analytics, and geospatial analysis. The firm's applied research spatial epidemiology, rural health, and AI ethics in research practice. Kevin A. Matthews, PhD, holds 27 years of experience in demographic applications of GIS. The policy monitor referenced throughout this paper is maintained at matthewsgeographics.com/policy-monitor.html.