

Tel Aviv University
Faculty of Social Sciences
Department of Sociology

Reconceptualising Authorship in the Age of Generative AI: A Theoretical Review

Academic paper submitted for the course:

Digital sociology

In partial fulfillment of the requirements for the Bachelor's degree in Social Sciences

Submitted by:

Adam Barak

July 2026

Table of Contents

Abstract	1
1. Introduction	2
2. Literature Review	4
2.1. Scope and Source Selection.....	4
2.2. Authorship as Spectrum: From Binary to Continuum.....	4
2.3. Legal and Institutional Frameworks Under Pressure	6
2.4. Tensions and Debates	7
3. Discussion / Theoretical Synthesis.....	9
3.1. Theoretical Synthesis.....	9
3.2. Position within the Field.....	11
3.3. Theoretical Contribution.....	12
4. Limitations	14
5. Future Research.....	15
6. Summary	16
6.1. Recap of the Argument.....	16
6.2. Concluding Remarks	17
References.....	18

Abstract

The emergence of generative AI writing tools has renewed long-standing debates about what it means to be an author. This paper argues that generative AI does not rupture but rather intensifies pre-existing theoretical instabilities in the concept of authorship, and that no single existing framework — legal-copyright, intentionalist-philosophical, or posthumanist-assemblage — is individually sufficient to account for AI-mediated academic writing.

The argument draws on legal theory, analytic epistemology, posthumanist scholarship, and the sociology of the text, synthesising these traditions to map their respective explanatory limits when applied to large language models (LLMs) in academic contexts. The paper moves across intentionalist, instrumentalist, and assemblage-based accounts to examine where each framework holds and where it breaks down.

Three central claims follow from this analysis: authorship has always been a contested, distributed concept rather than a stable individual property; human-AI collaboration in writing is best understood as a transparency-regulated continuum of epistemic delegation rather than a binary of authorship versus ghost-writing; and academic integrity frameworks require reconceptualisation around epistemic agency and disclosure rather than originality alone. Together, these claims reframe authorship attribution as a governance question as much as a philosophical one, with direct consequences for how institutions approach intellectual property, co-authorship credit, and scholarly accountability in an era of generative AI.

Keywords: generative AI, authorship attribution, epistemic agency, human-AI collaboration, academic integrity, large language models (LLMs), ghost-writing, intellectual property

1. Introduction

Authorship has never been a settled concept. Long before generative AI entered academic writing, scholars across literary theory, philosophy of mind, and intellectual property law were already contesting what it means to be the originator of a text — who counts as an author, what cognitive or creative labour that status requires, and what institutional interests its attribution serves. The arrival of large language models (LLMs) capable of producing coherent, discipline-specific prose has not created this instability; it has made it impossible to ignore.

The specific problem motivating this paper is the inadequacy of the frameworks currently deployed to adjudicate AI-assisted academic writing. Legal copyright doctrine rests on categories — originality, creative labour, and legal personhood — that are structurally anthropocentric and cannot be extended to non-human generative processes without doctrinal incoherence. Intentionalist philosophy of authorship, which locates meaning and responsibility in the mental states of a human agent, struggles to account for outputs whose generative process involves no intention in any recognisable sense. Posthumanist and assemblage-theory perspectives, while better equipped to accommodate distributed and non-human agency, risk dissolving authorship so thoroughly that scholarly accountability becomes unworkable. Each framework illuminates something the others obscure, yet none is individually sufficient. The result is a conceptual vacuum at precisely the moment when institutions need actionable criteria for academic integrity.

This paper addresses that vacuum by asking how competing theoretical frameworks conceptualise authorship when generative AI tools mediate academic writing. The aim is not to declare one framework correct, but to examine what each can and cannot explain, and to argue that a provisional synthesis — one that reconceives authorship as a transparency-regulated continuum of human–AI delegation — represents the most analytically coherent path forward for scholarly accountability. Five supporting claims structure this argument: that authorship was already theoretically unstable before generative AI emerged; that the binary opposition between AI-as-tool and AI-as-autonomous-agent is analytically untenable; that existing copyright law is structurally ill-equipped to govern AI-assisted texts; that originality must be reconceptualised as a relational rather than intrinsic property; and that transparency of AI participation, rather than origin-based originality, should serve as the operative standard for academic integrity.

Situating these claims requires recognising that debates about authorship have deep roots. Behzadi et al. (2015) demonstrate, through analysis of pre-modern Arabic literary culture, that

authorship has historically encompassed precursory, executive, collaborative, and revisionary functions — a reminder that the solitary, originating author is a comparatively recent and culturally specific ideal rather than a universal baseline. As that analysis observes, "to identify authorship as a form of human work is to validate individual agency" (Behzadi et al., 2015) — a formulation that reveals how much normative weight the concept carries beyond its descriptive function. The stakes of that weight become visible in contemporary academic contexts, where recognition for intellectual labour remains, as Knöchelmann (2023) documents across a survey of more than 1,000 humanities scholars and social scientists, "a central struggle" shaped by institutional pressures and national research frameworks. When the origin of a text is genuinely ambiguous — as it is whenever a researcher delegates drafting, structuring, or synthesis to a generative AI system — the institutional machinery built around authorial recognition begins to malfunction in ways that no single existing framework can repair.

The paper is organised as follows. Section 2 reviews the relevant theoretical literature across legal-copyright, intentionalist, and posthumanist traditions, tracing how each has handled the question of non-human or distributed textual production. Section 3 synthesises those strands into a critical discussion that tests the central argument against the frameworks' respective strengths and limits. Section 4 identifies the principal limitations of the analysis, and Section 5 considers directions for further inquiry. Section 6 provides a summary of the paper's core argument and its implications for academic integrity.

2. Literature Review

2.1. Scope and Source Selection

This review draws on peer-reviewed scholarship published between 2015 and 2025, spanning literary theory, philosophy of language, legal studies, science and technology studies, and the emerging field of AI governance. The guiding question — how competing theoretical frameworks conceptualise authorship when generative AI tools mediate academic writing — required sources capable of addressing authorship not merely as a technical or administrative category but as a contested theoretical object. The corpus was therefore assembled to include poststructuralist and posthumanist theoretical work, copyright and intellectual property scholarship, empirical studies of how readers and writers respond to AI-generated texts, and governance analyses of AI systems in institutional contexts. Sources were selected for their direct relevance to at least one of three analytical levels: the philosophical status of the authoring subject, the legal architecture that regulates authorship attribution, and the institutional norms — particularly those governing academic integrity — that operationalise authorship in practice.

A deliberate methodological choice shapes what follows: rather than proceeding source by source, this review organises its findings thematically, tracing lines of agreement, productive tension, and unresolved disagreement across the corpus. This approach suits a theoretical review whose central argument is that no single framework is individually adequate. Where sources share a claim, the convergence is noted; where they diverge, the disagreement is examined on its own terms rather than smoothed into a false consensus. Two broad thematic axes structure the discussion — the reconceptualisation of authorship as a spectrum rather than a binary, and the legal and institutional pressures that AI-assisted writing places on existing frameworks — before the review turns to the tensions and debates the corpus leaves genuinely open.

2.2. Authorship as Spectrum: From Binary to Continuum

The most persistent finding across the theoretical literature is that the binary opposition between human author and non-human tool is analytically untenable. This is not a conclusion that generative AI has imposed on the field; it is one that poststructuralist literary theory reached decades earlier and that AI-assisted writing has now made practically urgent. Henrickson (2021) famously dissolved the author into a tissue of citations and cultural codes, arguing that the unity

of a text lies not in its origin but in its destination — the reader (Barthes, 2017). Rodrigues (2025) extended this analysis by proposing the author-function — the idea that 'author' names a regulatory discourse that classifies, distributes, and controls texts rather than a person who creates them (Foucault, 2019). What generative AI adds to this theoretical inheritance is not a new argument but a new empirical pressure: when a large language model produces a paragraph that a human researcher then revises, selects, and submits, the question of what the author-function is regulating becomes practically, not merely philosophically, difficult to answer.

This pressure is most sharply articulated in scholarship that examines how authorship is assigned rather than how it is exercised. Fathallah (2017) demonstrates, in the context of fanfiction communities, that authorial identity is actively constructed through what she terms the 'legitimation paradox': communities simultaneously invoke and contest the authority structures of professional authorship in order to achieve cultural recognition. Fanfiction writers claim authorial status by appealing to the same criteria of originality, creativity, and voice that their practice structurally undermines. The parallel with AI-assisted academic writing is instructive. Researchers who use generative AI tools to draft, revise, or synthesise material face a structurally identical paradox: they must assert authorial accountability within institutional frameworks premised on individual, originary creation, while their actual practice distributes the generative work across human and algorithmic processes. The legitimation paradox, in other words, is not a curiosity of fan culture; it is a structural feature of any authorial practice that operates at the margins of a dominant framework.

This reconceptualisation of authorship as a continuum — a spectrum of agency, intentionality, and delegation rather than a categorical distinction — gains further support from posthumanist scholarship, which frames authorial agency as an emergent property of assemblages rather than a capacity of isolated individuals. On this account, the human researcher who uses a large language model is not the sole author of the resulting text, but neither is the model: authorship is distributed across intentions, prompts, training data, institutional constraints, and the specific choices made in revision and selection. The practical implication is that authorial accountability cannot be located at a single point; it must be tracked across the entire process of text production. This is a theoretically coherent position, though it raises a governance challenge the posthumanist literature has not yet resolved: if agency is genuinely distributed, how should academic integrity

policies assign responsibility when the resulting text is plagiarised, fabricated, or otherwise deficient?

2.3. Legal and Institutional Frameworks Under Pressure

If the theoretical literature converges on the view that authorship is a spectrum, the legal and institutional literature converges on a different and more troubling finding: existing frameworks were not designed for spectra. Copyright law, in most jurisdictions, requires a human author and an original creative act; it allocates rights and responsibilities to identifiable legal persons, not to processes or assemblages. The emergence of generative AI exposes the anthropocentric assumptions embedded in these frameworks with unusual clarity (Fenwick & Jurcys, 2023). When an AI system generates a text that is statistically novel — in the sense that it has not appeared verbatim in any training document — but is not the product of human intention in the conventional sense, copyright doctrine has no settled answer to the question of ownership. Different jurisdictions have responded with different provisional answers, none of them theoretically satisfying: some deny copyright protection to AI-generated works entirely, some attribute it to the human who operated the system, and some leave the question open pending legislative reform.

The governance dimension of this problem extends well beyond copyright. Birkstedt et al. (2023) find, in a systematic review of AI governance scholarship, that the field remains fragmented and characterised by a lack of explicit definitions, with limited understanding of how governance principles are actually implemented in practice. This fragmentation matters for academic integrity because universities and publishers are attempting to formulate disclosure and attribution policies in the absence of any settled governance framework. The result is a proliferation of institutional responses — some prohibiting AI use entirely, some requiring disclosure, some treating AI as a permitted tool analogous to a grammar checker — that reflect not considered policy but the absence of one. The analogy to earlier technological transitions (word processing, search engines, reference management software) is frequently invoked to suggest that the disruption is manageable, but it is weaker than it appears: those technologies mediated the expression of ideas the human author had already formed, whereas generative AI participates in the formation of the ideas themselves, or at least in the generation of the textual forms through which ideas are expressed and evaluated.

This distinction — between tools that assist expression and tools that participate in ideation — is central to the question of epistemic agency. If a researcher uses a large language model to generate a literature review section and then incorporates it with minimal revision, the epistemic work of identifying, evaluating, and synthesising sources has been at least partially delegated to the model. The researcher's name on the paper implies a form of intellectual accountability that the actual process of production may not support. Transparency and disclosure requirements address the surface of this problem — they make the delegation visible — but they do not resolve the deeper question of whether disclosed AI participation is compatible with the forms of scholarly accountability that authorship attribution is designed to guarantee.

2.4. Tensions and Debates

The corpus does not resolve into consensus, and the most analytically productive move is to hold its central tensions open rather than to adjudicate them prematurely. The sharpest disagreement concerns whether AI language systems possess any form of authorial agency or are strictly non-agential instruments. One position, grounded in phenomenological and sociological analysis, holds that readers behaviourally assign agency to AI-generated outputs — that the author-function, as a regulatory discourse, is applied to texts regardless of whether a human being produced them. The competing position, grounded in philosophy of mind and epistemology, holds that large language models lack genuine intentionality and that epistemic agency remains exclusively human. These two claims are not directly contradictory: the first describes how authorship is attributed in practice, while the second describes the metaphysical status of the systems producing the text. Both can be true simultaneously, which is precisely why the tension is generative rather than resolvable. The practical implication is that institutional frameworks must account for the gap between how authorship is perceived and how it is actually constituted — a gap that disclosure requirements alone cannot close.

A second unresolved tension concerns the appropriate normative response. Transparency-based positions — requiring researchers to disclose AI participation and specify its scope — are institutionally tractable and compatible with existing academic integrity and copyright infrastructures. Radical deconstruction positions, which argue that the individual-author model is itself the problem and should be replaced by collaborative or processual alternatives, are theoretically richer but have not been implemented at scale in mainstream academic publishing. The transparency approach is better supported for the immediate practical context, but it risks

preserving a framework whose conceptual foundations have been undermined. A third tension concerns the historical stakes of the present moment: is AI-assisted writing a rupture in the history of authorship, or an intensification of instabilities that were already present? The historical plurality of authorship functions — collaborative, revisionary, oral, executive — suggests the latter, but the speed and scale at which generative AI can produce text that passes as human-authored introduces qualitative pressures that historical analogy alone cannot fully absorb. Taken together, these tensions define the theoretical terrain that the remainder of this paper navigates.

3. Discussion / Theoretical Synthesis

3.1. Theoretical Synthesis

The preceding sections traced three distinct theoretical trajectories through which scholars have attempted to account for authorship in AI-mediated academic writing: the legal-copyright tradition, which anchors authorship in human intentionality and proprietary originality; the intentionalist-philosophical tradition, which locates authorial meaning in the cognitive commitments of a human agent; and the posthumanist-assemblage tradition, which redistributes agency across material, technological, and institutional actors. What this review has shown is not that generative AI invalidates any one of these frameworks, but rather that it exposes the internal tensions each framework was already carrying before large language models entered academic writing practice. The central argument of this paper — that generative AI intensifies pre-existing theoretical instabilities rather than inaugurating an entirely new crisis — finds its clearest support in the way each framework strains at precisely the same point: the question of where propositional and epistemic responsibility resides when text is produced through human-AI collaboration.

That convergence of strain is analytically significant. The legal tradition struggles because copyright doctrine was designed to protect expressions of human creativity, and its requirement of human authorship becomes procedurally unstable when the generative process is substantially delegated to a large language model. The intentionalist tradition struggles because intentionalism presupposes a unified cognitive agent whose purposes determine textual meaning, and that presupposition is disrupted when argumentative structure, disciplinary claims, and even rhetorical choices may originate in a statistical model trained on prior human texts. The posthumanist tradition struggles in the opposite direction: by distributing agency so thoroughly across the assemblage of human, algorithm, training data, and institutional context, it risks dissolving the very locus of accountability that academic integrity requires. Each framework captures something real, and each, taken alone, is insufficient. The synthesis this paper proposes — reconceiving authorship as a transparency-regulated continuum of human-AI delegation — does not claim to have resolved these tensions theoretically; rather, it argues that the tensions can be managed

institutionally through disclosure practices that keep a human agent accountable for the epistemic commitments embedded in a text.

The question of whether large language models themselves possess a form of authorial agency is one this synthesis deliberately sidesteps at the ontological level, because the practical question for academic integrity governance is not whether LLMs have genuine intentionality but who bears responsibility for the claims a text makes. Bajohr (2024) addresses this indirectly through the concept of 'causal authorship' — a metric of the distance between human and machine agents in the creative process — which usefully reframes the question from one of metaphysical status to one of causal and epistemic proximity. When a human writer supplies the research question, evaluates the output, revises the argument, and takes responsibility for the final claims, the causal distance between human intention and textual product is relatively short, even if a large language model generated substantial portions of the prose. When a writer submits an AI-generated text with minimal engagement, that distance widens to a point where the human's epistemic contribution becomes genuinely difficult to identify. The continuum this paper proposes is therefore not a binary between 'human author' and 'AI author' but a graduated spectrum along which disclosure requirements calibrate to the degree of delegation.

Zeilinger's posthumanist agential assemblage framework complicates this picture productively. By rejecting the anthropomorphic view of 'creative AI' and insisting instead that AI art practices 'provoke a necessary reconfiguration of the author function by entangling aesthetics with law and capital' (Zeilinger, 2021), Zeilinger captures something the causal-distance model risks missing: the institutional and economic forces that shape which configurations of human-AI collaboration are recognised as legitimate authorship at all. Academic publishing is not a neutral infrastructure; it embeds norms of originality, individual attribution, and proprietary knowledge that have historically served particular interests. Adema (2021) approaches this from a different angle, arguing that scholarly publishing is constrained by the commodification of print and that authorship should shift toward 'anonymous collaborative authorship' and 'processual' publications that embrace dynamic materiality over static originality. The force of this critique is real: the individual-author model is not a neutral description of how knowledge is actually produced but a regulatory and economic convention. The normative conclusion that the model should therefore be dismantled does not follow straightforwardly from that descriptive critique, however — particularly in the context of academic integrity, where individual accountability for epistemic

claims serves a function beyond intellectual property. It is the mechanism by which scholarly communities assign responsibility for error, fabrication, and misrepresentation.

3.2. Position within the Field

This paper's argument occupies a specific and somewhat uncomfortable position between two dominant camps in the current literature. On one side, scholars such as Rodrigues (2025) and Israel and Amer argue that the appropriate response to AI-mediated writing is to strengthen transparency and disclosure requirements within existing authorship frameworks, preserving human epistemic agency as the governing norm. On the other side, Adema and Zeilinger argue that the individual-author model is itself the theoretical and institutional problem, and that generative AI provides an occasion to pursue more radically collaborative, processual, and non-proprietary conceptions of scholarly production. This paper's synthesis does not simply split the difference. It accepts the posthumanist critique as analytically productive — the assemblage description of how AI-assisted texts are actually produced is more accurate than the fiction of the solitary human author — while rejecting the normative conclusion that accountability can or should be distributed to the point of unassignability.

The transparency-and-delegation framework developed here is also in dialogue with Israel & Amer (2022), whose analysis of AI-generated content argues that the emergence of synthetic text 'necessitates a re-evaluation of traditional mechanisms of attribution and the nature of authorship in digital scholarship.' Their proposal for a technical infrastructure capable of tracking data provenance and ensuring accountability — the MetaScribe architecture — represents one institutional instantiation of the transparency principle this paper defends at the theoretical level. The approach is architecturally ambitious and practically unproven at scale, but it shares with this paper's synthesis the core commitment that the appropriate response to distributed text production is not to abandon attribution but to make its conditions more explicitly traceable.

The objection that generative AI is merely the latest in a long sequence of writing technologies — and that the authorship concept has accommodated the printing press, the typewriter, spell-checkers, and grammar assistants without fundamental reconceptualisation — deserves a direct response. This paper concedes that authorship has always been technologically mediated and that historical alarmism about new writing tools has frequently proved disproportionate. The categorical difference, however, lies in what large language models actually produce: not formatting assistance, surface-level correction, or information retrieval, but propositional content,

argumentative structure, and disciplinary claims — elements that have traditionally been constitutive of intellectual authorship itself. When a tool generates the claim rather than merely conveys it, the threshold of epistemic delegation has crossed into territory that prior frameworks were not designed to govern. This is not a rupture, but it is a qualitative escalation that requires theoretical revision even if it does not require wholesale abandonment of existing concepts.

3.3. Theoretical Contribution

The principal contribution of this review is synthetic rather than originary. By placing the legal-copyright, intentionalist-philosophical, and posthumanist-assemblage frameworks in direct comparative dialogue — and by tracing the point at which each encounters the same underlying problem — the review maps the theoretical terrain in a way that makes visible both what each tradition offers and where each falls short. Bajohr's 'distributed authorship' concept (Bajohr, 2024), Zeilinger's posthumanist agential assemblage (Zeilinger, 2021), Adema's processual and collaborative publishing model (Adema, 2021), and Israel and Amer's provenance-tracking infrastructure (Israel & Amer, 2022) are not competing answers to the same question so much as partial responses to different dimensions of the same problem. The synthesis this paper proposes treats them as complementary rather than mutually exclusive: a distributed description of text production (Bajohr; Zeilinger) is compatible with concentrated accountability for epistemic commitments (Rodrigues; Israel and Amer), provided that the human author's degree of delegation is transparently disclosed.

That framing carries a specific implication for academic integrity governance. If authorship is reconceived as a continuum of human-AI delegation rather than a binary human-versus-machine distinction, institutional policies can be calibrated to the degree of delegation rather than attempting to draw a bright line between permissible tool use and impermissible ghost-writing. The ghost-writing analogy, invoked by several commentators to characterise unrestricted AI use in academic writing, is instructive precisely because it reveals what is at stake: not the origin of the words, but the authenticity of the epistemic commitment they express. A student who submits an AI-generated argument as their own has not merely used a tool; they have delegated the intellectual labour that academic assessment is designed to evaluate. The transparency-regulated continuum addresses this by requiring disclosure of the degree and nature of AI involvement, thereby preserving the evaluative function of academic authorship attribution without demanding an impossible return to the fiction of unmediated individual composition.

What this synthesis does not resolve should be stated plainly. The frameworks synthesised here are drawn overwhelmingly from Western literary theory, Anglo-American and Continental European legal traditions, and a relatively small number of empirical studies. The transparency-and-delegation model is offered as the most analytically coherent position available within this corpus, not as a universal prescription. The absence of non-Western, Global South, and multilingual academic writing contexts from the available scholarship represents a genuine limitation, and the synthesis should be read as a provisional theoretical position — one that identifies the most defensible path forward given the current state of the literature, while acknowledging that the path may look substantially different when viewed from outside the traditions on which it draws.

4. Limitations

Several limitations bearing on the scope and conclusions of this review warrant acknowledgement. Most fundamentally, this paper is entirely theoretical in orientation: it synthesises and critically compares existing frameworks rather than generating or testing empirical claims of its own. Consequently, the argument that a transparency-regulated continuum offers the most analytically coherent path forward remains a reasoned synthesis rather than a demonstrated finding — its practical validity depends on empirical work that lies beyond this paper's remit.

A second constraint concerns the disciplinary range of the literature reviewed. The evidence base draws predominantly on literary theory, philosophy of language, and legal studies, with comparatively limited engagement with empirical research in educational psychology or higher-education pedagogy. This skew means that the experiential dimension of human-AI collaboration — how students and academic staff actually navigate, perceive, and rationalise AI-mediated writing in practice — is addressed only obliquely. The frameworks analysed speak to what authorship conceptually is or should be, but they engage less directly with what it feels like or functionally does within institutional settings.

Third, the literature year range of 2015 to 2025 captures a field in rapid transition. Scholarship on generative AI and academic writing is accumulating at a pace that makes any review liable to partial obsolescence; policy positions and institutional norms that appear emergent here may already have shifted in the time between drafting and reading. As a theoretical review, this paper also cannot adjudicate between competing frameworks on empirical grounds alone: the choice of synthesis over any single framework ultimately reflects an interpretive judgement, and scholars working from different disciplinary commitments may reasonably weigh the evidence differently.

5. Future Research

Several directions emerge from the limitations of existing scholarship that warrant sustained empirical and theoretical attention. Most urgently, the frameworks examined throughout this paper — legal-copyright, intentionalist, and posthumanist-assemblage — have been developed and debated almost exclusively through conceptual analysis; empirical investigation into how authors, readers, and institutional gatekeepers actually negotiate questions of epistemic agency and accountability in AI-mediated writing contexts remains sparse. Longitudinal studies tracking how academics' practices and self-perceptions of authorship attribution shift as generative AI tools become more deeply embedded in research workflows would substantially enrich the largely theoretical picture this paper has assembled.

A second productive avenue concerns disciplinary variation. The instabilities in authorship that this paper has traced may manifest quite differently across fields with distinct norms of attribution, collaboration, and originality — the sciences, humanities, and professional disciplines each carry different assumptions about what intellectual contribution looks like. Comparative empirical work across these contexts could test whether a transparency-regulated continuum of human-AI collaboration, as provisionally proposed in the Discussion, operates as a genuinely cross-disciplinary model or requires significant domain-specific calibration.

The relationship between institutional policy and actual academic behaviour also deserves closer examination. Whether disclosure requirements, as currently formulated by universities and journals, meaningfully alter authors' sense of intellectual responsibility — or merely redistribute accountability onto procedural compliance — remains an open question. Research combining policy analysis with qualitative methods such as interviews, think-aloud protocols, or discourse analysis of submitted work could illuminate the gap between normative frameworks and lived scholarly practice.

6. Summary

6.1. Recap of the Argument

This paper set out to examine how competing theoretical frameworks conceptualise authorship when generative AI tools mediate academic writing. The central argument advanced throughout has been that generative AI does not introduce an entirely unprecedented rupture into the concept of authorship; rather, it intensifies and renders newly visible instabilities that were already latent within existing frameworks. Three broad traditions were brought into dialogue: the legal-copyright tradition, which conditions authorship on the demonstration of human creative originality and therefore struggles to accommodate outputs produced through human-AI collaboration; the intentionalist-philosophical tradition, which locates authorship in purposive mental states and encounters difficulty when those states are distributed across human prompting and machine generation; and the posthumanist-assemblage tradition, which dissolves the unified author-subject into a network of human and non-human actants and thereby offers the most theoretically permissive account of AI-mediated textual production, yet risks evacuating the normative grounding that academic integrity requires.

None of these frameworks proved individually sufficient. The legal tradition provides institutional clarity, but purchases it at the cost of conceptual flexibility, leaving a growing domain of human-AI co-authored work in a state of regulatory ambiguity. The intentionalist tradition preserves the moral weight of authorial responsibility yet depends on a model of cognitive sovereignty that posthumanist critique has substantially eroded. The posthumanist tradition, for its part, captures the genuinely distributed character of AI-mediated writing while offering little guidance on the accountability structures that institutions and scholarly communities must maintain. The synthesis proposed in the discussion section — reconceiving authorship as a transparency-regulated continuum of human-AI delegation — does not resolve these tensions definitively, but it provides a more analytically coherent basis for navigating them than any single framework can offer alone. This position draws indirect support from scholarship that has itself proposed continuum-based thinking: Henrickson (2021), working from empirical reader-response data, argued that a spectrum running from authorship to generatorship better

captures the graduated reality of computer-generated text than any binary classification, a logic that transfers readily to the institutional questions surrounding academic writing.

6.2. Concluding Remarks

The practical and theoretical stakes of this inquiry are considerable. If academic institutions continue to adjudicate questions of authorship and academic integrity using frameworks built for a world of solitary human writers, they risk both conceptual incoherence and practical injustice — penalising forms of human-AI collaboration that are functionally indistinguishable from other long-tolerated forms of intellectual assistance, while simultaneously failing to address the more substantive epistemic agency concerns that underlie integrity norms in the first place. The contribution of this paper lies not in settling the debate but in mapping its fault lines with sufficient precision to make productive institutional and scholarly responses more tractable.

The deliberate boundary of this work is its restriction to theoretical and conceptual sources; no empirical data on how writers, readers, or institutions actually experience or negotiate AI-mediated authorship in practice were gathered or analysed. That restriction leaves the synthesis intact as a theoretical account, but it also marks the point at which the argument reaches its natural limit. The most pressing extension of this work lies in moving from conceptual mapping to situated, context-sensitive study — examining how the transparency-regulated continuum model holds up when tested against the lived practices of academic writers operating under real institutional constraints, rather than against the theoretical positions those institutions have yet to fully articulate.

References

- Adema, J. (2021). Living Books. *The MIT Press eBooks*.
<https://doi.org/10.7551/mitpress/11297.001.0001>
- Bajohr, H. (2024). Writing at a Distance: Notes on Authorship and Artificial Intelligence. *German Studies Review*, 47(2), 315–337. <https://doi.org/10.1353/gsr.2024.a927862>
- Barthes, R. (2017). The Death of the Author. *The Continental Aesthetics Reader*.
<https://doi.org/10.4324/9781351226387-35>
- Behzadi, L., Hämeen-Anttila, J., Ghersetti, A., Gutas, D., & Görke, A. (2015). Concepts of authorship in pre-modern Arabic texts. *Bamberger Orientstudien*.
<https://doi.org/10.20378/irb-40118>
- Birkstedt, T., Minkkinen, M., Tandon, A., & Mäntymäki, M. (2023). AI governance: themes, knowledge gaps and future agendas. *Internet Research*, 33(7), 133–167.
<https://doi.org/10.1108/intr-01-2022-0042>
- Fathallah, J. M. (2017). Fanfiction and the Author. *Amsterdam University Press eBooks*.
<https://doi.org/10.1515/9789048529087>
- Fenwick, M., & Jurcys, P. (2023). Originality and the future of copyright in an age of generative AI. *Computer Law & Security Review*, 51, 105892.
<https://doi.org/10.1016/j.clsr.2023.105892>
- Foucault, M. (2019). What Is an Author? *Textual Strategies*.
<https://doi.org/10.7591/9781501743429-006>
- Henrickson, L. (2021). Towards a new sociology of the text: the hermeneutics of algorithmic authorship. In *Figshare*. <https://doi.org/10.26174/thesis.lboro.10596404.v1> [\[verify details\]](#)
- Israel, M. J., & Amer, A. (2022). Rethinking data infrastructure and its ethical implications in the face of automated digital content generation. *AI and Ethics*, 3(2), 427–439.
<https://doi.org/10.1007/s43681-022-00169-1>
- Knöchelmann, M. (2023). Authorship and Publishing in the Humanities. In *Cambridge University Press eBooks*. Cambridge University Press.
<https://doi.org/10.1017/9781009223089>

Rodrigues, T. V. (2025). Distant writing and the epistemology of authorship: On creativity, delegation, and plagiarism in the age of AI. *International Journal of Social Sciences and Humanities Invention*, 12(5), 8598–8613. <https://doi.org/10.18535/ijsshi/v12i05.04>

Zeilinger, M. (2021). Tactical entanglements: AI art, creative agency, and the limits of intellectual property. In *Social Science Open Access Repository (GESIS – Leibniz Institute for the Social Sciences)*. <https://doi.org/10.14619/1839> *[verify details]*