



The Ghost in the Machine: AI-Generated Literature and the Reconfiguration of Creative Authenticity

Dr Shipra Malik

How to Cite this Article:

Malik, D. S. (2026). The Ghost in the Machine: AI-Generated Literature and the Reconfiguration of Creative Authenticity. International Journal of Creative and Open Research in Engineering and Management, 2(7), 1-9.
<https://doi.org/10.55041/ijcope.v2i7.220>

License:

This article is published under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

© The Author(s). Published by International Journal of Creative and Open Research in Engineering and Management.



<https://doi.org/10.55041/ijcope.v2i7.220>

Abstract

For most of literary history, the question of who wrote a text carried an assumption so obvious it rarely needed stating: a person did. The fluency of contemporary large language models has unsettled that assumption, producing sonnets, short stories, and novel chapters that readers in controlled studies frequently cannot distinguish from human-authored work. This paper argues that AI-generated literature does not so much destroy literary authenticity as force a reconfiguration of what the term has always meant. Synthesizing literary theory, empirical reception research, copyright law, and consumer psychology, the paper traces four converging lines of evidence: readers judged blind often cannot detect machine authorship and sometimes rate it more favorably, yet the same readers penalize a text emotionally once its AI origin is disclosed; structural analyses show AI-generated fiction and poetry remain measurably less inventive than the best human work even when formally polished; and copyright authorities have sidestepped the detection problem entirely by anchoring protection to traceable human decision-making rather than textual quality. Revisiting Roland Barthes's "death of the author" alongside a 2016 case in which an AI-assisted novella nearly won a Japanese literary prize, the paper argues that authenticity was never a property readers detected inside a text but a social practice negotiated among readers, critics, publishers, and legal institutions. It concludes that a more defensible, procedural model of authenticity is

emerging one grounded in disclosed, traceable human judgment rather than the simple fact of a byline.

Keywords : AI-generated literature; authorship; authenticity; large language models; Roland Barthes; computational creativity; copyright law; literary reception studies

Introduction

Attribution has always been the quiet precondition of literary meaning. For most of literary history, the question "who wrote this?" carried an assumption so obvious it rarely needed stating: a person did. That assumption underwrote everything readers built on top of a text the emotional trust extended to a memoir, the premium paid for a signed first edition, the prestige a national prize conferred on a debut novelist. Authenticity, in this sense, functioned less as an aesthetic judgment than as a form of currency, one whose value depended entirely on a stable, if often invisible, guarantee: that a traceable human consciousness sat behind the words on the page. That guarantee is no longer safe. Contemporary large language models can now produce sonnets, short stories, and full-length novel chapters on request, and in controlled studies, ordinary readers frequently cannot tell these outputs apart from work by celebrated human poets and novelists (Porter and Machery; Marklová et al.). This is not a marginal technical curiosity confined to novelty demonstrations; it is a direct challenge to the interpretive and institutional scaffolding that literary culture has built around the idea of the author.

The stakes of this reconfiguration extend well beyond academic debate. Authenticity has long functioned as a kind of currency in the literary marketplace: it underwrites the emotional trust a reader extends to a memoir, the premium a collector pays for a signed first edition, and the prestige a national prize confers on a debut novelist. When the origin of a text becomes uncertain or deliberately concealed, that currency destabilizes, and every institution built on top of it, from the publishing contract to the plagiarism policy, must renegotiate its terms. This paper proceeds in several movements: first, a brief history of how authorship became tied to authenticity in the first place; second, a plain account



of what large language models technically do when they "write"; third, a review of empirical studies testing whether readers can detect machine authorship; fourth, an examination of the specific creative limits current models still exhibit; fifth, a look at how copyright law has drawn its own, quite different line around human authorship; and finally, a discussion of what a more distributed, process-oriented model of authenticity might look like once the old byline-based guarantee can no longer be trusted at face value.

Authorship Before the Algorithm: The Romantic Inheritance

The notion that a text's value depends on the singular consciousness behind it is a comparatively recent invention, tied to the Romantic image of the solitary genius whose inner feeling guarantees the truth of the work. Twentieth-century literary theory complicated this picture considerably. Roland Barthes's influential essay on the death of the author argued that a text's meaning is generated in the act of reading rather than fixed by the biography or intention of whoever composed it, and that treating the author as the ultimate source of meaning artificially closes off interpretation (Barthes 142-48). Barthes was not writing about machines, but his argument anticipates a problem that generative AI now makes literal: if meaning was never entirely anchored to a knowable, intentional human mind in the first place, what exactly is lost when that mind is replaced by a statistical model trained on other people's sentences? Scholars revisiting Barthes in the context of AI have suggested that large language models extend, rather than invent, this dispersal of authorial control, since the model itself synthesizes patterns from thousands of previous authors rather than originating meaning from a single point of view (Filimowicz). The historical irony is that a theory once used to dethrone the human author is now recruited to explain why a text with no human author at all can still function as literature for a reader who has no way of verifying its origin.

Machines That Write: What Large Language Models Actually Do

It is worth being precise about the mechanism behind "AI-generated literature," because the term invites both overstatement and dismissal. Contemporary large language models generate text by predicting, one token at a time, which word or word-fragment is statistically most plausible given everything that precedes it, having learned these probabilities from enormous archives of previously published human writing. There is no episodic memory of a childhood, no lived grief or joy motivating a line of verse, and no communicative intention in the way a human poet intends to move a particular reader. What the model produces is best described as a highly sophisticated recombination of stylistic and narrative patterns already present in its training data, arranged in response to a prompt. This is precisely why debates over AI creativity so often return to the distinction, proposed by cognitive scientist Margaret Boden, between combinational creativity, which recombines existing ideas in new juxtapositions, exploratory creativity, which works within an established conceptual space, and transformational creativity, which alters the rules of that space entirely (Boden). Current-generation LLMs perform combinational and, to a lesser degree, exploratory creativity fluently; empirical research suggests they still fall short of the transformational leaps associated with major human literary innovation, tending toward formal correctness and stylistic conformity rather than genuine departure (Wilkins). The "ghost in the machine," then, is not a hidden consciousness but a compressed statistical portrait of a very large number of dead and living authors, reanimated on demand.

Empirical Turing Tests: Can Readers Tell the Difference?

If the philosophical case against AI authorship rested on readers being able to detect the absence of a human mind, that case is empirically weaker than it might seem. A widely cited 2024 study published in *Scientific Reports* had participants judge poems by canonical human poets alongside AI-generated imitations and found that readers were not reliably able to identify which was which; strikingly, participants on average rated the AI-generated poems more favourably than the human originals, apparently because the machine output tended to be more accessible and less ambiguous (Porter and Machery). A follow-up line of research extended this question beyond English, testing Czech readers on a mix of human- and AI-authored poetry in a language that makes up a tiny fraction of most models' training data. Participants again performed close to chance when guessing authorship, and the study additionally found an "authorship bias": once a reader believed a poem was AI-generated, they rated it less favourably, even when, unbeknownst to them, they had judged the very same poem more highly moments before under the belief it was human-written (Marklová et al.). Together these findings point to something important: authenticity, in practice, functions less as a property that readers detect in the text itself and more as a label that reorganizes their evaluation after the fact. The text does not change; the meaning of "authentic" attached to it does.



This labelling effect is not confined to poetry readers judging art for its own sake. Consumer research on AI-authored marketing messages found that when audiences believed an emotionally expressive communication had been written by an AI rather than a person, they reported reduced positive word of mouth and weaker loyalty toward the brand responsible, an effect the researchers traced to a drop in perceived authenticity and a corresponding rise in moral discomfort (Kirk and Givi). Notably, this "AI-authorship effect" nearly disappeared when the communication was purely factual rather than emotional, and it reversed when audiences learned the message had simply been copied from elsewhere rather than freshly composed by a person (Kirk and Givi). The implication for literature is direct: readers do not object to AI output because it is worse by any measurable formal standard; they object because emotionally resonant writing is culturally coded as evidence of a human interior, and learning that no such interior was involved retroactively cheapens the emotional transaction the reader believed they were having.

The Creativity Question and the Limits of Statistical Imagination

None of this means AI-generated literature is creatively equivalent to human-authored literature in every respect, and the more rigorous studies are careful to say so. Comparative analyses of AI-generated verse against large corpora of human poetry have found that machine output, while formally competent, gravitates toward a narrow set of default choices: a preference for rhyme and quatrain structure, a heavy reliance on first-person-plural address, and a comparatively restricted vocabulary relative to the range found across human poets (Wilkens). Similar work comparing AI-generated short stories to work by professional authors, evaluated using structured creativity assessments, has found that machine-written fiction passes markedly fewer tests of narrative originality, and that AI-generated plots tend to recycle a smaller set of recurring structures across repeated generations, in contrast to the greater structural diversity typical of human-written plots over the same sample size (Wilkens). This pattern is consistent with Boden's framework: a system trained to predict the statistically likely next word will, almost definitionally, be pulled toward the centre of its training distribution rather than its outer edges, which is where transformational creative breakthroughs tend to occur. The "reconfiguration" this paper's title points to, then, is not a levelling of human and machine creativity but a narrowing of the gap in some registers (technical competence, surface fluency, formal correctness) alongside a persistence of the gap in others (structural novelty, thematic risk, sustained originality across a large body of work).

Whose Hand Guided the Pen? Copyright Law and the Bedrock of Human Authorship

The law has had to confront the authenticity question in a much blunter form than literary theory: not "does this feel authentic," but "who, if anyone, owns it." In 2023 the United States Copyright Office issued formal guidance stating that copyright protection requires human authorship, and that applicants seeking to register a work containing more than a trivial amount of AI-generated material must disclose which portions were AI-generated and describe the human contribution involved (United States, Copyright Office, "Copyright Registration Guidance"). The Office subsequently refused registration in several test cases, including a graphic novel whose AI-generated illustrations were denied protection even though the human-authored text and the overall selection and arrangement of the work were registrable, and an artwork whose applicant had failed to disclose the AI-generated portions as required (United States, Copyright Office, "Copyright Registration Guidance"). A more detailed 2025 report reaffirmed this position while clarifying that human contributions perceptible within an AI-assisted output, such as creative arrangement, selection, or meaningful modification of AI-generated material, remain eligible for protection even when the AI-generated substrate itself is not (United States, Copyright Office, Copyright and Artificial Intelligence). A federal court reviewing an appeal in this area went further, stating plainly that copyright law protects only the products of human creativity and that it has never been stretched to cover works generated by technology operating without any guiding human hand (United States, Copyright Office, Copyright and Artificial Intelligence). What emerges from this body of policy is a legal definition of authenticity that has nothing to do with whether a reader can detect a difference and everything to do with whether a traceable human decision shaped the expressive content. Law and aesthetics, in other words, have arrived at strikingly different tests for the same underlying question.

The Death of the Author, Reborn: Distributed and Posthuman Authorship

If neither empirical indistinguishability nor legal doctrine settles the authenticity question cleanly, a third framework worth taking seriously treats AI-assisted writing as a form of distributed authorship rather than a binary choice between "human" and "machine." On this view, a poem generated with the assistance of a language model is never solely the product of the model; it also carries the imprint of the engineers who built the system, the enormous and largely uncredited population of authors whose writing constitutes its training data, and the human user who selected a prompt,



curated an output, and decided what counted as a finished text. Writers who revisit Barthes's argument in light of generative AI note that this arrangement does not so much kill the author a second time as multiply the author into a network of contributors whose individual traces have become difficult to isolate (Filimowicz). This reframing has practical value: rather than asking whether an AI-generated poem is "authentic" in some absolute sense, it invites readers and critics to ask what kind and degree of human judgment shaped a given text, a question that admits of gradations rather than a simple yes or no. A human writer who spends an afternoon prompting, discarding, and revising machine-generated drafts until a single stanza survives intact has exercised a meaningfully different kind of authorship than one who copies an unedited first draft wholesale, even though a reader encountering only the finished page may be unable to tell which process produced it.

The Integrity Problem: Detecting, Disclosing, and Defending Human Writing

The difficulty readers have in distinguishing AI-generated from human-generated text has produced a small but growing research effort aimed at verifying authorship independent of the finished product. One proposed framework argues that because AI detection tools which analyze a finished text are increasingly unreliable against more sophisticated models, institutions concerned with academic and literary integrity should instead monitor the writing process itself, capturing the drafting, revising, and pausing behavior characteristic of human composition as a kind of behavioral signature distinct from anything a generative system produces in a single pass (Aburass and Rumman). Such process-based verification represents a tacit admission that the text alone, however closely examined, may no longer reliably answer the authenticity question; the evidence of human involvement has to be relocated to the history of the writing act rather than the artifact it produced. This shift has consequences for how publishers, literary prizes, and educational institutions will likely need to operate going forward, moving away from after-the-fact detection and toward documented, disclosed composition practices.

Case in Point: When a Computer Nearly Won a Literary Prize

The stakes of this reconfiguration became concrete a full decade before today's most capable models existed. In 2016, a short work titled "The Day a Computer Writes a Novel," produced through a collaboration between a research team at Future University Hakodate led by computer scientist Hitoshi Matsubara and a text-generation program the team had built, was submitted to Japan's Nikkei Hoshi Shinichi Literary Award, a prize that had previously opened its blind-judged submission process to non-human applicants ("An AI-Written Novella"). The entry passed the first round of screening among 1,450 submissions, and a judging novelist described it afterward as surprisingly well structured, though still weak in areas such as character description ("An AI-Written Novella"). The human team had set the plot parameters and the character genders in advance, with the program itself selecting the specific words and sentences within those constraints, making the finished text a genuinely mixed product rather than a purely autonomous one. This early episode anticipates, in miniature, every tension this paper has traced: a blind judging process that could not detect the machine's involvement, a human framework of intention wrapped around an algorithmic execution, and a literary institution forced to decide, in real time, what "authorship" would mean once it could no longer be assumed to belong entirely to a person.

Implications for Publishing, Education, and the Literary Marketplace

These findings are not confined to the laboratory; they are already reshaping how literary institutions behave. Publishers now routinely request disclosure of AI assistance in submitted manuscripts, a practice that mirrors the disclosure requirement the Copyright Office imposes on registration applicants, and several literary magazines have adopted explicit policies rejecting undisclosed AI-generated submissions outright, reasoning that the reader's trust in the publication depends on knowing what kind of process produced the work being read. Creative-writing programs face a related but distinct problem: if human evaluators cannot reliably distinguish AI-assisted student work from unassisted work, and if process-monitoring tools of the kind proposed by researchers of writing integrity are not yet standard practice in workshop settings, then the pedagogical value of the traditional creative-writing exercise, learning to generate original material under one's own effort, is quietly undermined even when no rule has technically been broken (Aburass and Rumman). Literary prizes face perhaps the sharpest version of this dilemma, since a prize is, among other things, a public certification of individual talent; a blind-judged competition that cannot detect machine involvement, as the 2016 Hoshi Shinichi case demonstrated, risks certifying achievement that no single person actually earned in the way audiences assume. None of these institutions has arrived at a settled answer. What they share is a recognition that the old proxy for authenticity, the simple fact of a byline, no longer carries the evidentiary weight it once did, and that new mechanisms of trust will have to be built to replace it.



Conclusion: Authenticity as Practice, Not Property

The evidence surveyed here resists a simple verdict on whether AI-generated literature is or is not authentic, because the studies examined answer different versions of that question. Readers, tested blind, frequently cannot detect the difference and sometimes prefer the machine's output (Porter and Machery; Marklová et al.), yet those same readers penalize a text emotionally once its AI origin is revealed (Kirk and Givi). Structural analysis suggests machine-generated fiction and poetry remain measurably less inventive than the best human work even when superficially fluent (Wilkens), while legal authorities have simply relocated the entire question to a test of traceable human decision-making rather than textual quality (United States, Copyright Office, Copyright and Artificial Intelligence). What unites these otherwise divergent findings is a shared conclusion: authenticity in literature was never a stable property waiting to be discovered inside a text, and the rise of generative AI has only made this fact impossible to ignore. It was always a social practice, an agreement among readers, critics, publishers, and legal institutions about which kinds of production count as sufficiently human to deserve the trust, admiration, or copyright protection that the label implies.

This does not mean the distinction between human and machine authorship is meaningless, only that it has to be relocated. The most defensible version of authenticity to emerge from this survey of evidence is neither purely textual nor purely legal but procedural: a text becomes authentic to the degree that a traceable, disclosed act of human judgment shaped its final form, whether that judgment consists of drafting every sentence unaided or of curating, revising, and selecting among machine-generated possibilities until something genuinely one's own remains. The ghost in the machine is not a false consciousness masquerading as a poet; it is the accumulated voice of thousands of real, uncredited human authors, refracted back through a system with no interior of its own. Whether that refracted voice deserves to be called literature is, in the end, a question literary culture will keep answering the same way it always has: not by inspecting the text for a hidden soul, but by deciding, collectively and often inconsistently, and always in the open rather than behind an undisclosed prompt, what it wants authorship to mean next.

Works Cited

Aburass, Sanad, and Maha Abu Rumman. "Authenticity in Authorship: The Writer's Integrity Framework for Verifying Human-Generated Text." arXiv, 16 Apr. 2024, arxiv.org/abs/2404.10781.

"An AI-Written Novella Almost Won a Literary Prize." Smithsonian Magazine, 25 Mar. 2016, www.smithsonianmag.com/smart-news/ai-written-novella-almost-won-literary-prize-180958577/.

Barthes, Roland. "The Death of the Author." *Image-Music-Text*, translated by Stephen Heath, Hill and Wang, 1977, pp. 142-48.

Boden, Margaret A. *The Creative Mind: Myths and Mechanisms*. 2nd ed., Routledge, 2004.

Filimowicz, Michael. "AI and the Rebirth of the Death of the Author: The Informational Advantages of Statistically Average Writing." Medium, Higher Neurons, 1 June 2026, medium.com/higher-neurons/ai-and-the-rebirth-of-the-death-of-the-author-the-informational-advantages-of-statistically-3777a277287e.

Kirk, Colleen P., and Julian Givi. "The AI-Authorship Effect: Understanding Authenticity, Moral Disgust, and Consumer Responses to AI-Generated Marketing Communications." *Journal of Business Research*, vol. 186, 2025. ScienceDirect, www.sciencedirect.com/science/article/abs/pii/S0148296324004880.

Marklová, Anna, et al. "The Author Is Dead, But What If They Never Lived? A Reception Experiment on Czech AI- and Human-Authored Poetry." arXiv, 26 Nov. 2025, arxiv.org/abs/2511.21629.

Porter, Brian, and Edouard Machery. "Artificial Intelligence versus Maya Angelou: Experimental Evidence That People Cannot Differentiate AI-Generated from Human-Written Poetry." *Scientific Reports*, vol. 14, 2024. Nature, www.nature.com/articles/s41598-024-76900-1.



United States, Copyright Office, Library of Congress. "Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence." Federal Register, vol. 88, 16 Mar. 2023, www.federalregister.gov/documents/2023/03/16/2023-05321/copyright-registration-guidance-works-containing-material-generated-by-artificial-intelligence.

United States, Copyright Office. Copyright and Artificial Intelligence, Part 2: Copyrightability. Jan. 2025, www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-2-Copyrightability-Report.pdf.

Wilkins, Matthew. "AI as a Tool for Simulation-Based Experiments in Literary Studies." arXiv, 1 June 2026, arxiv.org/abs/2606.02293.