Indian Journal of Modern Research and Reviews

This Journal is a member of the *'Committee on Publication Ethics'*

Online ISSN:2584-184X

Research Article

Algorithmic Bias and AI-Generated Literature: Rethinking Authorship, Intention, and Critical Hermeneutics in the Age of AI

Mohd Younis Bhat [1*], Dr. Aashaq Hussain Zargar [2], Dr. Sheweta Magotra [3]

[1] Assistant Professor, Department of English, Government Degree College Tral, Higher Education Department, Jammu and Kashmir, India

[2] Assistant Professor, Department of English, Government Degree College Tral, Higher Education Department, Jammu and Kashmir, India

[3] Assistant Professor, Department of English, Desh Bhagat University, Punjab, India

Corresponding Author: * Mohd Younis Bhat

DOI: https://doi.org/10.5281/zenodo.24099689

Abstract

The proliferation of Large Language Models (LLMs) and deep generative neural networks (DGN) in creative writing, across a wide range of domains, has disrupted the current paradigms of literary production and critical theory. This paper challenges the intersection of algorithmic bias and synthetic literature by showing how autoregressive architectures can perpetuate, amplify and naturalize biases found in web-scale training corpora, thereby propagating historical biases. This study looks at the survivability of the concept of artistic intentionality and critical hermeneutics in a method of statistical token prediction of literary text, using the theoretical frameworks of Roland Barthes's post-structuralist critique of the author, Michel Foucault's author-function and N. Katherine Hayles' distributed cognition. Reflecting on recent interventions in Critical (AI) studies (2023-2026), the paper explores computational creativity, positing that contemporary generation systems have reduced the semantic insight in favor of syntactic fluidity, and have compelled cross-entropy-led linguistic flattening rendering non-Western epistemologies, vernaculars and marginalised idioms vulnerable to disenfranchisement. Finally, the paper proposes a post-humanist critical discourse, which is neither technologically deterministic nor reactionary humanist; it sees literature as an act of emotional, historical and socio-political resistance.

Manuscript Information

- **ISSN No:** 2584-184X
- **Received:** 17-09-2026
- **Accepted:** 06-11-2026
- **Published:** 10-11-2026
- **MRR:**4(11); 2026: 72-76
- **©2026, All Rights Reserved**
- **Plagiarism Checked:** Yes
- **Peer Review Process:** Yes

How to Cite this Article

Bhat MY, Zargar AH, Magotra S. Algorithmic Bias and AI-Generated Literature: Rethinking Authorship, Intention, And Critical Hermeneutics in the Age of AI. Indian J Mod Res Rev. 2026;4(11):72-76.

Access this Article Online

www.mrrjournal.in

Keywords: Algorithmic Bias, AI-Generated Literature, Algorithmic Authorship, Intentionality, Author-Function, Posthumanism, Computational Creativity, Linguistic Flattening.

© 2026 Mohd Younis Bhat, Dr. Aashaq Hussain Zargar, Dr. Sheweta Magotra. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY NC ND).https://creativecommons.org/licenses/by/4.0/

1. INTRODUCTION

Literary production has been for a long time an apotheosis of the exceptionalism and subjective interiority of human beings. The Romantic conviction that poetry is "the spontaneous overflow of powerful feelings" coming from what Wordsworth calls "the contemplative faculty of the mind" (126); the Modernist use of stream-of-consciousness to reflect the disjointed human subjectivity; the author's role as a bearer of moral and historical truth have all traditionally depended on the author's witness. The texts were seen to be a product of lived experience: bereavement, desire, violence of structure, physical weakness, and collective memory of culture.

This consensus has been undone by the launch of transformer-based Large Language Models (LLMs) such as the GPT architectures from OpenAI, Claude from Anthropic, and Gemini from Google. These generative engines create formally valid sonnets, elaborate dramatic dialogues, elegies and recursive narrative prose in seconds, and pose a profound and pressing challenge for literary hermeneutics: Where is the aesthetic agency in an elegy generated by a matrix of tokens that self-referentially generates its form and through a criterion of probabilistic cross-entropy reduction?

These are computational ideologies that need to be critically interrogated in a transitioning process. Literary studies, as Rita Raley and Russell Samolsky suggest in their description of the emerging imperatives of Critical AI, must go beyond being superficially "fascinated by computational fluency" and engage in the fragmentation of the material, institutional, and political infrastructures that govern artificial intelligence (185–88). Generative models are not only trained on uncurated, web-scale corpora that reflect historical, cultural, and ideologic hierarchies, but they also function as such. As a result, synthetic literature exhibits an aesthetic paradox, namely that of an apparently inexhaustible formal novelty lying over a fundamental structural repetition of the linguistic past. Starting from literary theory, critical code studies and post humanist criticism, this paper examines algorithmic literature, suggesting that algorithmic bias in synthetic literature is not a transient technical issue, but rather an epistemological and aesthetic property of machine learning architectures; that the author-function is not entirely dissolved but rather became a corporatized, distributed network; and that the imperatives of stochasticity in computational text production lead to flattening of the aesthetic, standardizing –for the better– cultural forms in detriment of subaltern and regional expression.

2. Bias and the Poetics of Algorithmic Literature

The usual understanding of algorithmic bias in computer science is an engineering deficiency, in practice some deviation from demographic parity or a model weight error. In the field of literary studies, however, bias is to be perceived as some sort of ideological and aesthetic constructions.

2.1 Training Data as Colonial Archive

Current autoregressive language modeling methods heavily use large-scale datasets that are scraped from the web, like Common Crawl, WebText, digital library repositories, and public discussion forums. The overwhelming orientation of these archives is towards the digital production of the Global North; industrialized, English-speaking populations in particular socio-economic categories. Jasmina Tacheva and Ramakrishna Ramaswamy show that generative AI resides in an“AI Empire” that functions by systematically stripping away cultural capital and reinforcing the digital neocolonial hierarchy (7–12). Using Spivak's well-known work, Epistemic Violence and Subaltern Silence (271-75), it is noticed that the pre-training corpus is a contemporary colonial archive. The statistical model recovers the power dynamics of the empire in its source texts as it attempts to build up the maximum semantic value of the text across billions of parameters.

The loss function that drives the model toward higher accuracy in reproducing the cross-entropy of the text encourages the model to choose, by statistical salience, typical syntactic structures, archetypal characters, and cultural narratives. Therefore, it is not uncommon that generative systems that represent non-western geographies and marginalized communities rely on exoticism, on the victimization without historical agency, or on paternalistic rescuing motifs. Corporate data-based homogeneity reinscribes the subaltern voice effectively.

2.2 Aesthetic Flattening and the Erasure of Ostranenie

Inventions in literature have happened through the violation of forms, through the experimentation with grammar and through defamiliarization. Ostranenie, or "estrangement" is for Viktor Shklovsky art's fundamental role in "disrupting" habitual perception, so that one can recover the sensation of life: Art exists that one may feel things, to make the stone 'stony' (12). Just as in music, literature cannot be easily commodified according to the standardized rules of language.

Transformer models are, on the other hand, inherently stochastic machines. These move to the target of the central distribution mechanism of token likelihoods, creating a beautiful regression to the average. As Ted Chiang explains, LLMs are a “blurry JPEG of the web" and modern generative text systems “are more like a blending of averaged linguistic artifacts instead of anything more creative. Machine learning is also working in "latent spaces" (42-46), which are based on statistical probabilities and naturally "squelch" idiosyncratic linguistic anomalies for the sake of coherence.

This mediation of structures is made manifest in the production chain of synthetic literature. The seed prompt from the human user then needs to go through corporate safety architecture, system prompts and reinforcement learning filters before hitting the stochastic transformer engine. In that engine, the multi-head attention and matrix multiplications result in the probability distributions of synthetic tokens being emitted. Lastly, the text is fed into a human reader who imposes interpretive meaning on the output. The forms of deviance from the norm of the corporate English--whether they are non-standard patois, indigenous idioms, or vernacular hybridities--are statistically punished, flattened or normalized as syntax

© 2026 Mohd Younis Bhat, Dr. Aashaq Hussain Zargar, Dr. Sheweta Magotra. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY NC ND).https://creativecommons.org/licenses/by/4.0/

errors. The model's aesthetic production is thereby formally conservative and can be easily traced back to a corporate reinforcement learning from human feedback (RLHF) approach, which aims to maximize consensus and minimize friction.

3. Deconstructing Algorithmic Authorship

An analytical move towards post-structuralist discussion of authorial agency is required in the production of synthetic text.

3.1 The Literalization of the"Death of the Author"

In his 1967 essay Roland Barthes declared: “Writing is the destruction of every voice and every point of origin: Writing is that neutral, composite, oblique space where our subject slips away, the negative where all identity is lost, starting with the very identity of the body writing” (142). Barthes has criticized "biographical positivism," and has claimed that a text is "a tissue of quotations drawn from the innumerable centers of culture"(146).

The Large language models are literally Barthes' thesis. The language model has neither subjective interiority, nor mortal body, nor individual trauma. It behaves as an echo chamber of intertextuality, decomposing cultural products created by humans into multi-dimensional vector embeddings and recombining tokens into sentences.

However, this literalization "poses an acute theoretical irony. Indeed, as David J. Gunkel suggests in revisiting Barthes and Foucault in the era of current AI developments, the death of the author in the generative system does not democratize literature, but rather moves discursive control to the corporate black box and the infrastructures of private platforms (124-28). The algorithmic death of the author takes textuality to private places, whereas Barthes had conceived of the death of the author as emancipatory empowering the reader and liberating textuality. Rather than an empowered reader allowing the decentralized author to be replaced, it is proprietary algorithms with weights and corporate alignment protocols that constitute the boundaries of expression.

3.2 Foucault and the Distribution of the Author-Function

As the author-function is defined by Michel Foucault, it is not a natural creative person but an apparatus that regulates, constricts and labels discourse in legal, economic and institutional structures (115-18).

In the world of generative AI, that author-function no longer refers to any one individual, but is spread throughout an asymmetrical network. In discussing legal and philosophical issues about prompt engineering and copyright, Francesca Mazzi notes that a collision between the two challenges the notion of attribution of creative agency (52-56). The author-function is distributed over four different kinds of nodes: the corpus contributors, who, by their unpaid and uncredited creative and academic work, provide the foundation for training; the corporate architects and aligners, who design the loss objectives, safety guardrails and curation filters; the prompter or interrogator, who provides the direction or iterated “lexical steering” to the probabilistic generation; and the computational engine itself, which performs the linear algebraic calculations over the weights of the parameters. If synthetic literature meets issues of intellectual property, copyright violation, aesthetic culpability, it is unable to find an autonomous subject in the author-function. Authoring is a corporate-enabled process of uncoupling discourse from accountability and focusing financial gain to platform monopolies.

4. The Ghost of Intentionality

Human literary creation and machine text generation is a meeting point in the phenomenological intentionality. This divergence takes place on ontological, linguistic, hermeneutic, and textual registers. From an ontological perspective, human literature has an embodied, mortal and situated consciousness, while algorithmic generation is purely based on mathematical calculation on high dimensional token matrices. Another important distinction between human and machine writing is that in the former case the semantics — the words which are connected to lived experience and social reality — are created while in the latter case the syntax — the manipulation of random symbols in accordance with statistical rules — is executed. Human literary expression is motivated by ethical, formal and socio-political purposes, algorithmic text – by prompt parameters, temperature settings and loss minimization (hermeneutic approach). Finally, while human writers work with textual tradition by means of formal break, intertextual subversion and historicized criticism, computational writers work by means of stochastic recombination, interpolation and regression to statistics.

4.1 Symbol Manipulation versus Phenomenological Meaning

John R. Searle's "Chinese Room “thought experiment has shown that systematic symbol manipulation does not result in semantic comprehension (417–19). When used on literature, the language model turns into an automated Chinese Room. The algorithm isn't necessarily in contact with mourning or happiness, systemic oppression or flesh decaying. Processes only the numeric distances between the vector embeddings.

The ability to create pathos for a human reader does not reflect machine empathy for a human reader when it comes to AI-generated poems. Instead, the written words serve as a mirror of emotions: the depth of emotions is all of the reader's own making. The reading of electronic literature and AI conversational interfaces is frequently a psychological projection of the reader and thus an interiorizing experience for the reader, as described by Scott Rettberg (8–11). The machine gives out signifiers, the reader projects the signified.

4.2 The Automated Intentional Fallacy

In New Criticism, W. K. Wimsatt Jr. and Monroe C. Beardsley cautioned against the "Intentional Fallacy" which meant that the author's side of the story in the process of writing a literary

© 2026 Mohd Younis Bhat, Dr. Aashaq Hussain Zargar, Dr. Sheweta Magotra. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY NC ND).https://creativecommons.org/licenses/by/4.0/

work should not be taken into account, but rather, what he had written should be examined (468, 470).

Synthetic literature brings with it another inverse phenomenon, the Automated Intentional Fallacy. Readers have been socialized by human communication to read artifacts of language with the assumption of purposive design, implication of subversive irony, and symbolicity. The generation of an unusual metaphor or juxtaposition often results in the reader thinking of it as an avant-garde artistic move, when in fact it is a normal part of LLM usage. From a practical perspective, the phrasing is simply a random sampling from a long tail of a probability distribution (a stochastic event caused by the temperature). Mathematical noise is a conscious stylistic project by the interpretive community.

5. Re-Evaluating Computational Creativity

To place the boundaries of algorithmic creativity into perspective, it will be useful to refer to Margaret A. Boden's taxonomy of creativity: Combinatorial creativity, which involves "combining familiar ideas in rather unusual ways" (3); Exploratory creativity, "which explores known conceptual areas while observing certain underlying rules" (4); and Transformational creativity, which involves "changing the conceptual space itself, by breaking its rules, rewriting them, or otherwise extending them" (6).

These large language models are great at combinatorial, exploratory activities. They sum up an Elizabethan sonnet in a legalistic way with one breath and write a detective story in a science fiction mode with another.

But transformational creativity is beyond the reach of a computer. LLMs are constrained by the scope of their weight and produce text in line with historical precedent. They do not change the form of the literature when stretched beyond their statistics, but go into incoherence or hallucination. There have been some important literary revolutions that have come out of neglectful refusal: refuse the stories told to you by the State, refuse the syntax of the colonizer, refuse structural forms. An aesthetic or political refusal cannot be simulated by a statistical engine.

6. Toward a Post-humanist Critical Approach

N. Katherine Hayles warns against both uncritical techno-utopianism and oppositional humanism in her foundational work in the fields of cybernetics and literature; she instead argues for an analytical model that is based on distributed cognition (1–5; Unthought 11–14). Hayles argues that cognition exists not only as the product of an independent Cartesian subject, but is as much a matter of human agents, biological organisms, code settings and technical platforms. For generative neural networks, Ajay Kumar Shukla and Nitesh Kumar Singh claim that the algorithmic writing signals a hybrid literary system, in which social memory, that of humans, merges with nonconscious computing activities (52-55).

A post-humanist literary criticism cannot go back to Romantic claims of an untarnished heart. Rather, researchers need to come up with critical methodologies that are attuned to human-machine writing spaces:

First of all, scholars need to develop strong algorithmic criticism and aesthetics of forensics. Literary theorists need to explore synthetic texts by teasing out the institutional, computational and architectural conditions that brought them about (Kirschenbaum 12–15), and reveal the dialects, cultural viewpoints, and political positions that have been left out in data curation and safety alignment.

Secondly, there is a need to reposition human writers and critics to reclaim the role of literature as embodied resistance. While automation moves towards smarter formulaic and commercial substances, human writers must accentuate what automation cannot mimic: geospecific, phenomenologically situated and politically vulnerable writing.

Finally, the discipline is required to confront monopolies of the corporate platforms. Algorithms' bias is solidified by private tech infrastructures and literary criticism should counter the monopolization of natural language. Scholars, writers and cultural institutions need to be the proponents of open-source architectures, community data sets that are in the interest of the public, and archives that are multilingual and not quintessentially Western.

7. CONCLUSION

Generative artificial intelligence is not a death knell for literature; rather, it represents a profound challenge to, and interrogation of, the fundamental social, ideological, and political contexts in which literature is created and understood. Algorithmic bias in synthetic data is not a minor glitch to be patched, but an inevitable consequence of training predictive neural networks on historically accumulated and culturally conditioned digital archives. The large language model represents the consensus of culture; when it produces writing, prevailing cultural conventions dominate its language and are presented as natural, as though its creativity were ideologically neutral.

Generative models indeed reveal the limitations of traditional notions of individual authorship and the absence of genuine intentionality in automated prose and poetry, forcing readers to confront a fundamental question: why read and write at all?

If literature is nothing more than syntactic dexterity, linguistic pattern-matching, and the masquerade of established genres, then machines have caught up with human writers. Yet, if literature remains, as Susan Sontag argued, a moral way of being—an act of human responsibility and a mortal, situated, existential encounter with vulnerability, political struggle, and the pursuit of liberation—then synthetic language models can never fully replace the human voice of literature. It is the responsibility of literary scholars, critics, and authors to keep literature uncomputable: rooted in collective memory, radical imagination, and enduring resistance.

REFERENCES

1. Bajohr H, editor. Thinking with AI: machine learning the humanities. Open Humanities Press; 2025.

© 2026 Mohd Younis Bhat, Dr. Aashaq Hussain Zargar, Dr. Sheweta Magotra. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY NC ND).https://creativecommons.org/licenses/by/4.0/

2. Barthes R. The death of the author. In: Image, music, text. Translated by Heath S. Fontana Press; 1977. p. 142–8.
3. Boden MA. The creative mind: myths and mechanisms. 2nd ed. Routledge; 2004.
4. Braidotti R. The posthuman. Polity Press; 2013.
5. Chiang T. ChatGPT is a blurry JPEG of the web. The New Yorker. 2023 Feb 9. Available from: https://www.newyorker.com/tech/annals-of-technology/chatgpt-is-a-blurry-jpeg
6. Foucault M. What is an author? In: Bouchard DF, editor. Language, counter-memory, practice: selected essays and interviews. Translated by Bouchard DF, Simon S. Cornell University Press; 1977. p. 113–38.
7. Gunkel DJ. Creating with AI: on recent debates about authorship—revisiting the influence of Barthes and Foucault. Stud Univ Babes-Bolyai Philos. 2024;69(2):121–38.
8. Hayles NK. How we became posthuman: virtual bodies in cybernetics, literature, and informatics. University of Chicago Press; 1999.
9. Hayles NK. Unthought: the power of the cognitive nonconscious. University of Chicago Press; 2017.
10. Kirschenbaum MG. Mechanisms: new media and the forensic imagination. MIT Press; 2008.
11. Mazzi F. Authorship in AI-generated works: exploring originality in text prompts and AI outputs through philosophical foundations of copyright. J World Intellect Prop. 2024;27(1):45–68.
12. Raley R, Samolsky R. Critical AI: a field in formation. Am Lit. 2023;95(2):185–204.
13. Rettberg S. Creative interactions with large language models: electronic literature and the performance of authorship. Humanities. 2025;14(1):1–17.
14. Searle JR. Minds, brains, and programs. Behav Brain Sci. 1980;3(3):417–24.
15. Shklovsky V. Art as technique. In: Russian formalist criticism: four essays. Translated by Lemon LT, Reis MJ. University of Nebraska Press; 1965. p. 3–24.
16. Shukla AK, Singh NK. Algorithmic authorship: AI and the changing nature of literary creation. English J. 2025;7(2):49–58.
17. Spivak GC. Can the subaltern speak? In: Nelson C, Grossberg L, editors. Marxism and the interpretation of culture. Macmillan; 1988. p. 271–313.
18. Tacheva J, Ramaswamy R. AI empire: unravelling the interlocking systems of oppression in generative AI's global order. Crit AI. 2023;1(1–2):1–28.
19. Wimsatt WK Jr, Beardsley MC. The intentional fallacy. Sewanee Rev. 1946;54(3):468–88.
20. Wordsworth W. Preface to Lyrical Ballads (1802). In: Owen WJB, Smyser JW, editors. The prose works of William Wordsworth. Vol. 1. Oxford University Press; 1974. p. 118–59.

Creative Commons License

This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution–Non-commercial–No Derivatives 4.0 International (CC BY-NC-ND 4.0) License. This license permits users to copy and redistribute the material in any medium or format for non-commercial purposes only, provided that appropriate credit is given to the original author(s) and the source. No modifications, adaptations, or derivative works are permitted.

About the Corresponding Author



Mohd Younis Bhat is an Assistant Professor in the Department of English at Government Degree College Tral, Higher Education Department, Jammu and Kashmir, India. His academic interests include English literature, literary criticism, cultural studies, language, and contemporary literary perspectives. He is actively engaged in teaching, academic research, and scholarly contributions.

© 2026 Mohd Younis Bhat, Dr. Aashaq Hussain Zargar, Dr. Sheweta Magotra. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY NC ND).https://creativecommons.org/licenses/by/4.0/