

Blind to the Machine: Reader Response Theory in an Age of Artificially Intelligent Authors

Satya Prakash Prasad, Abhishek Jha & Shafaque Zehra

Abstract

This research paper explores the implications of AI-generated literature for reader response theory and literary criticism. With advanced AI able to produce human-quality creative writing, assumptions about authorial intent and meaning rooted in human subjectivity are disrupted. The article argues reader response theory offers a model for interpretation severed from traditional notions of authorship. It also examines how schools of criticism reliant on authorial biography require new perspectives when analysing AI-authored texts. Overall, the emergence of machine creativity compels reinvention but also reconnection with humanism.

Keywords: Artificial intelligence, creativity, reader response theory, authorship, literary criticism

Introduction

In recent years, rapid advances in artificial intelligence have brought systems capable of human-level creativity and artistry tantalizingly close to reality. From painting startlingly vivid original images and composing intricate musical compositions to penning literature good enough to fool even discerning critics and editors, AI is beginning to match and potentially surpass human gifts for self-expression and imagination. For instance, AI models like Anthropic's Claude, Google's PaLM system and OpenAI's GPT-3 can engage in thoughtful philosophical dialogues, explain complex scientific concepts in clear prose, and even churn out poems, stories, and articles which are coherent and meaningful, if not

yet reaching the pinnacles of literary achievement (Bommasani et al., 2021; Bender et al., 2021). While much attention focuses, understandably, on the technological prowess underpinning these intelligent systems, their emergence raises crucial questions for the arts and humanities as well. As audiences increasingly encounter films, music, visual art, poetry, and fiction generated by AI algorithms which are more and more indistinguishable in form and style from works produced by gifted human creators, how might our understanding of artistic expression itself adapt and evolve in response? What new perspectives on meaning-making and evaluation become necessary when machine learning rather than individual human genius gives rise to cultural artefacts?

Reader response theory, pioneered by theorists like Wolfgang Iser, Louise Rosenblatt, and Stanley Fish in the mid-twentieth century, offers a useful framework for examining these questions. At its core, reader response theory assumes that a literary work takes place in the mutual relationship between the reader and the text wherein the reader is “not seen as a separate entity, acting upon the environment, nor the environment acting on the organism, but both parts acting as a total event” (Rosenblatt, 1978, p. 98). While radical when first introduced, subsequent postmodern and post structural conceptions of blurring authorship align well with the reader response perspective. Some even speak of the metaphorical “death of the Author”—the rejection of the Author-God in favour of more diffused questions of influence, interpretation, and intertextuality. Yet living human authors have always persisted even through postmodernism’s height, their shadowy presence still shaping assumptions about inspiration, intent and meaning making. Presumably, AI creative systems render even these lingering modernist assumptions obsolete.

Roland Barthes’ famous essay *The Death of the Author* (1977) declares audaciously that a text consists of multiple writings woven from innumerable cultural influences rather than springing fully realized from one godlike Author figure. As artificial intelligence systems advance in their literary capabilities, the concept of authorship for AI-generated texts is set to expand beyond what even Barthes envisaged. The traditional notion of a single author giving rise to a work will give way to a more diffuse sense of authorship, with the AI model’s training data, neural network architecture, and computational methods all contributing to the text’s creation. Though produced by machines, these synthetic

compositions may read as eloquently human. AI authorship will compel us to reimagine authorial voice as an emergent phenomenon borne of code, data, and statistical learning algorithms.

This research paper seeks to explore the implications of increasingly sophisticated AI literature through the lens of reader response theory. As artificial intelligence matches and starts exceeding human literacy, memory and emotional intelligence, the fundamental subjectivity of meaning creation through dynamic reading gains renewed importance. Reader response concepts provide a model for understanding text fundamentally severed from traditional notions of authorship and genesis. Rather than viewing AI's ability to write novels or poetry as diminishing humanity's artistic primacy, a reader-response perspective allows us to appreciate this technology for finally allowing art to be explored entirely through readers' imagination, unconstrained by the director's shadow of a human authorial figure.

Reader Response Theory

The origins of reader response theory are generally traced to Wolfgang Iser's *The Implied Reader* (1974), which emphasises reading as an interaction between reader and text over time. He contends that literary texts contain gaps and indeterminacies that require active reader participation in "concretization" to bridge missing connections and arrive at meaning (p. 134). An entire arena of interpretation unfolds in this process, shaped profoundly by readers' subjective experiences and changing responses as they move through a text. Iser conceives of reading as a dialectic dance between original perspective and shifts in synthesis, with literary works designed not to achieve one perfect reading but to spur readers' imaginations in an ongoing process of sense-making.

In parallel to Iser, Louise Rosenblatt also centrally influenced reader response theory through her transactional theory of reading articulated in *Literature as Exploration* (1938) and *The Reader, the Text, the Poem* (1978). She agrees with Iser that meaning emerges from the reading process itself. However, her transactional perspective emphasises how both reader and text transform in that process, stating that the literary work exists in a "live circuit set up between 'reader and text'" (Rosenblatt, 1978, p. 14). This live circuit shapes the meaning created each time a reader interacts with a text, making interpretation specific to particular readings rather than contained only in the text itself.

Expanding from Iser's and Rosenblatt's approaches, Stanley Fish's affective stylistics also informed reader response theory by exploring how rhetorical and linguistic devices elicit reader reactions. Fish (1970) argues that literary texts control interpretive choices through their formal construction, intentionally prompting emotional and sensory effects that orient readers' meaning making. However, Fish (1970) ultimately views readers themselves as paramount in the journey from device to interpretation, writing: "It is the reader who 'makes' literature" (p. 87). Though texts wield influence, true meaning remains contingent on readers' activity.

The fundamental principles underlying reader-response theory emphasize reader participation and the rejection of authorial intent. According to this perspective, texts lack inherent meaning; instead, significance emerges dynamically as readers interact with the symbols encoded in writing during the reading process. The interpretation of a text is subject to variations stemming from readers' diverse experiences, attitudes, expectations, and evolving mental states as they engage with and emotionally process the textual content. Reader-response theory posits that readers play an active and creative role in constructing meaning from literary works, asserting that without their participation, texts remain static artefacts devoid of defined interpretation or cultural value. Iser (1989) refers to text as a location which "is made up of a world that is yet to be identified and is adumbrated in such a way as to invite picturing and eventual interpretation by the reader" (p. 250).

Reader Response Theory in the Age of AI

The advent of creative AI that can generate original literary and artistic works has profound implications for reader response theory. When machines can write books and poems without any human inspiration or intent behind them, the reader truly is cut off from making meaning based on assumed authorial design. This forces the reader to derive meaning entirely from the text itself and its interaction with the reader's own experiences and interpretive lenses.

Some scholars argue that creative works generated by AI algorithms still inherently contain traces of human authorship via the original programming, training data, and design of the AI system (Wingström et al., 2022; Hutson, 2024). However, the exponential advancement of

deep learning techniques and the use of unsupervised learning by AI platforms increasingly removes the hand of human engineers. The raw text generated by such AI emerges more organically from statistical associations learned by the system through ingesting massive textual data sets, not from specific human-authored code or content (Pieper, 2020; Floridi & Chiriatti, 2020; Dale, 2021).

This growing rift between human authorship and AI-generated text resurrects Barthes' rallying cry against fixation on authorial intent. When AI models like GPT-3 produce a novel or poem, no original human author exists from which to derive meaning. The reader cannot interpret motifs, symbols, or ideological perspectives in the work by connecting them to an author's background or intended message. This liberates the reader to engage directly with the textual artefact itself.

Reader response theory unpacks how different readers' subjective experiences shape their interpretation of literary works. Iser (1978) described reading as an active process where the reader fills "gaps" in the text using imagination to construct meaning (p. 40). Fish (1980) focused on how readers connect a text to familiar cultural conventions and "interpretive communities" to derive meaning. Such reader-centred analysis is vital when assessing AI-generated creative output, given the total absence of an intelligible authorial presence.

Making sense of these machine-generated narratives cannot rely on divine inferences about authorial intent. The reader must work bottom-up from the content of the text itself and filter it through their subjective knowledge and assumptions. Making this interpretive leap free from assumed human origins behind the writing represents the most unfettered test case of reader response theory.

Creative AI also invites questions about the reader's ethical obligations when interpreting such works. If an AI system generates disturbing or problematic text, does the reader have a duty to correct this by carefully framing their response, given the lack of authorial agency? Literary scholar Seán Burke (2008) discussed the "aesthetic or ethical criteria" needed when choosing "between incompatible literary readings" (p. 205). Perhaps similar care is needed when analyzing AI output, so as not to carelessly propagate harmful themes that emerged non-intentionally from an unguided machine learning process.

While reader response theory argues that meaning lies with readers

rather than authors, Umberto Eco's concept of the "model reader" shows how authors guide interpretation through intentional narrative signals in the text itself (Eco, 1979, pp. 7-11). AI-authored works lack such inherent cues, leaving readers radically unbound. This unprecedented freedom demands responsible reading and thoughtful filtering of AI content through social and ethical lenses before determining meaning.

The uncontrolled randomness of AI-generated text also complicates reader response analysis. If an AI system outputs completely nonsensical passages, can the reader construct any coherent meaning at all? The concept of "wandering viewpoint" suggests that meaning emerges from the reader piecing together fragmentary phrases (Iser, 1974, p. 87). But at some point of incoherence, the reader may disengage, unable to posit even a subjective interpretation. Indicating when AI text crosses this threshold into unreadable chaos is key for reader response theorists assessing creative AI output. Also, elimination of authorial intent as a source of meaning in such cases creates a need for updated models of the reading process and new reader response methodologies when confronting works produced wholly without human intervention.

The Cyborg Author: Blurring Boundaries in AI-Generated Literature

The emergence of AI-generated literature not only challenges traditional notions of authorial intent but also blurs the boundaries between human and machine creativity. Donna Haraway's influential essay "A Cyborg Manifesto" (1985) provides a useful framework for understanding the implications of this hybrid authorship. Haraway defines the cyborg as a "cybernetic organism, a hybrid of machine and organism, a creature of social reality as well as a creature of fiction" (p. 5). In the context of AI-generated literature, the cyborg author can be seen as a fusion of human and machine agency, where the AI system's outputs are shaped by the human-designed algorithms, training data, and interaction with human readers.

The notion of the cyborg author challenges the humanist assumption of a singular, unified authorial consciousness behind a text. Instead, it suggests a distributed and collaborative form of authorship that emerges from the complex interplay between human and machine intelligence. As Haraway argues, the cyborg is a "creature in a post-gender world" (p. 8), signalling a move beyond binary oppositions and hierarchies.

Similarly, AI-generated literature invites us to rethink the binary between human and machine creativity, acknowledging the ways in which they are inextricably entangled.

Moreover, the cyborg author problematizes the idea of originality and authenticity in literary creation. If a text is generated by an AI system that has been trained on a vast corpus of human-authored works, to what extent can it be considered original or authentic? The cyborg author embodies what Haraway (1985) calls “the utopian dream of the hope for a monstrous world without gender” (p. 181), a world in which the boundaries between original and derivative, human and machine, are radically destabilized.

In this context, reader response theory becomes even more crucial for understanding the meaning and value of AI-generated texts. When the author is a cyborg, a hybrid of human and machine agency, the reader’s role in constructing meaning becomes paramount. The reader must grapple with the uncanny, disorienting experience of engaging with a text that is both familiar and alien, recognizably human-like in its language and style, yet also the product of inhuman, algorithmic processes.

Ultimately, the figure of the cyborg author invites us to embrace the fluidity and multiplicity of literary creation in the age of AI. It challenges us to move beyond the humanist obsession with individual genius and originality, and instead to recognize the ways in which human and machine creativity are always already entangled. By doing so, we can develop new, post humanist models of authorship and interpretation that are better suited to the complexities of our technological age.

Implications for Literary Theory and Criticism

The emergence of advanced AI models capable of generating literature with minimal or even without any human intervention has the potential to fundamentally disrupt several schools of literary theory and criticism. The exploration of such literary works created wholly outside of human subjectivity undermines the core tenets of major critical frameworks. This extreme test case of AI authorship pushes literary theory into new territory regarding interpretation, meaning, and the very nature of literature itself.

Since the days of Aristotle, a fundamental aspect of literary criticism

has been the examination of creative works through the authorial intentions. Formalists like Cleanth Brooks aimed to excavate organic unity in texts by viewing them as external manifestations of authors' inner experiences (Leake, 2021, p. 16). Psychoanalysts like Sigmund Freud and Jacques Lacan used biographical context to dissect authors' unconscious desires encoded symbolically in writing (Abrams, 1999, pp. 247-249). Postcolonial thinkers like Edward Said and Gayatri Spivak highlight how authors' cultural positions shape the colonial discourse in their works (Ashcroft et al., 1989).

AI-generated literature ruptures these intrinsic ties between author biography and textual meaning. The presence of an AI author severs the text entirely from notions of specific human creative intent or vision. When machines compose novels or poems through pattern analysis of immense textual data, no singular authorial consciousness exists from which to derive definitive meaning. The symbolic connections, textual motifs, ideological perspectives, and other literary elements which human readers have been trained to infer clues from about authorial intentions no longer provide an available avenue to make sense of the work.

This rupture fully liberates the reader from any cues or guidance about interpretation embedded deliberately by a human writer. The reader must work bottom-up from the content of the text itself, filtering it through their own subjective knowledge and assumptions to construct meaning. Even the concept of intertextuality, commonly invoked in postmodern theory to denote how creative works appropriate and transform elements of prior texts and culture is sure to be obscured in cases wherein AI models synthesise and recombine patterns from the training corpus in ways not neatly traceable to identifiable prior sources.

In essence, the possible emergence of a substantial body of literature independent of human authors compels an overhaul of major schools of literary theory while affirming the enduring necessity of human critical faculties. Assumptions embedded for centuries are untenable for AI-generated texts, requiring new perspectives on meaning and ethics. Reconceptualizing criticism and interpretation for this age of machine creativity is both a challenge and an opportunity to reinvigorate literary scholarship.

Analysing AI-Generated Literature: A Case Study

To illustrate the interpretive challenges posed by AI-generated literature, let us consider a sonnet (in Shakespearean style) generated by Open AI's GPT-3.5,

*In shadows cast by twilight's gentle hue,
Where whispers of the night dance softly by,
A poet's pen shall paint the skies anew,
With words that lift the spirit to the sky.
Oh, muse, inspire this humble bard once more,
To weave a sonnet with thy graceful touch,
Where dreams take flight on wings of lore,
And hearts are moved by love's enchanting clutch.
In every line, let passion brightly gleam,
Like stars that shimmer in the midnight air,
And let the rhythm flow like a gentle stream,
A melody that soothes away despair.
So here I pen this verse with heartfelt plea,
That in these words, true beauty all may see.*

(OpenAI, 2024).

This sonnet adheres to the traditional form and structure of the genre, with 14 lines, iambic pentameter, and a rhyme scheme of ABAB CDCD EFEF GG. The content of the poem, however, raises questions about the nature of creativity and inspiration in the context of AI-generated literature. The speaker invokes the "muse" and refers to the act of writing as a "poet's pen shall paint the skies anew," suggesting a romantic notion of artistic inspiration and creation. However, the fact that this poem was generated by an AI system challenges these traditional ideas of authorship and creative genius.

When interpreting this sonnet, the reader must grapple with the absence of a human author's intentions, emotions, and experiences. The themes of beauty, love, and the power of words to "lift the spirit" and "soothe away despair" are present in the text, but they cannot be directly attributed to a singular authorial consciousness. Instead, the reader must construct meaning from an amalgamation of formal elements of the poem and their own subjective response to the language and imagery.

The Rise of Machine Learning: An AI-Led Literary Movement

The prospect of advanced AI models generating entire literary movements poses intriguing questions for the evolution of literary theory and criticism. If neural network models were to author a cluster of thematically linked texts considered profoundly innovative and influential, how might critics incorporate this machine-led “school” into conceptualisations of influence, periodization, and categorization? Thoughtfully, analysing such an AI-generated avant-garde highlights unexamined assumptions in current theoretical frameworks while gesturing towards more inclusive, progressive models.

Of course, attributing discrete “movements” to AI systems risks simplistic anthropomorphisation. The training corpora and technical architecture underlying neural networks shape their literary output more than intentional circulation of ideas. However, shared selection pressures and training data could theoretically lead clusters of AI models to exhibit coherent emergent styles (McGuffie & Newhouse, 2020). If generative algorithms consistently produce works deemed groundbreaking by human readers, this could constitute an AI-driven avant-garde tradition which challenges longstanding notions of literary movements emerging from shared manifestos, collaborations, and polemics between human writers.

Periodization schemes tracking successive literary eras might also require recalibration to incorporate AI-authored movements. Models like Ian Watt’s rise of formal realism or M.H. Abrams’ romantic/classical/modernist paradigms assume continuity between human-created genres (Watt, 1957; Abrams, 1953). Even postmodernism’s ruptures are framed as reactions against prior schools. Integrating a wholly AI-driven movement would strain this impulse towards continuity and dialectical evolution. The sudden flourishing of a machine-written style without discernible roots in prior human movements might constitute a profound break from traditional conceptualizations of literary history.

The ideal approach is to avoid simplistic anthropomorphising of algorithmic processes while still granting machine-written works respectful attention on their own terms. AI-created texts should be treated as rich artefacts for analysis, not hollow derivatives of “real” human literature. Tracing intertextual connections across both AI and human-written corpora can highlight their symbiotic creative potential

while avoiding reductive binaries.

Humanistic inquiry always evolves in response to cultural and technological shifts. Rather than defensively resisting algorithmic creativity, literary scholars should lead the way in developing novel frameworks for understanding AI's revolutionary aesthetic possibilities without forfeiting rigorous humanistic analysis. This synthesis promises to uphold ethics and meaning amidst digital upheaval.

Conclusion

The emergence of advanced artificial intelligence capable of generating human-quality literature compels a fundamental re-evaluation of core assumptions in literary theory and criticism. As AI algorithms increasingly author creative works without direct human involvement, notions of meaning rooted in authorial intent and subjectivity become disrupted. Reader response theory, which locates the construction of literary meaning in the interaction between text and reader, offers a valuable framework for interpreting and analyzing AI-generated literature. By emphasising the reader's active role in shaping interpretation, reader response perspectives allow for the appreciation of machine-authored works on their own terms, unconstrained by traditional humanist conceptions of inspiration and creative genius.

However, the rise of AI authorship also poses challenges for literary criticism, particularly schools of thought that rely heavily on authorial biography and historical context to excavate textual meaning. Scholars must develop new analytical tools and adapt existing models to thoughtfully engage with the unique qualities of AI-generated literature. This includes reconceptualising notions of literary movements, influence, and periodization to account for the potentially transformative impact of machine creativity. Ultimately, the advent of AI literature presents an opportunity for a more inclusive, dynamic understanding of the nature of art and meaning-making—one that celebrates the irreducible diversity of human imagination while also making space for new forms of creative expression made possible by technological innovation. By embracing this synthesis of humanistic inquiry and computational creativity, literary theory can evolve to meet the challenges and possibilities of an AI-driven future.

References

- Abrams, M. H. (1999). *A glossary of literary terms* (7th ed.). Earl McPeck.
- Abrams, M. H. (1953). *The mirror and the lamp: Romantic theory and the critical tradition*. Oxford University Press.
- Ashcroft, B., Griffiths, G., & Tiffin, H. (1989). *The empire writes back*. Routledge.
- Barthes, R. (1977). *Image, music, text* (S. Heath, Trans.). Fontana.
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big?. *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*. <https://doi.org/10.1145/3442188.3445922>
- Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demszky, et al. (2021). On the opportunities and risks of foundation models (Version 3). *arXiv*. <https://doi.org/10.48550/ARXIV.2108.07258>
- Burke, S. (2008). *The death and return of the author: Criticism and subjectivity in Barthes, Foucault, and Derrida*. Edinburgh University Press.
- Dale, R. (2021). GPT-3: What's it good for? *Natural Language Engineering*, 27(1), 113–118. <https://doi.org/10.1017/S1351324920000601>
- Eco, U. (1979). *The role of the reader: Explorations in the semiotics of texts*. Indiana University Press.
- Fish, S.E. (1970). Literature in the reader: Affective stylistics. *New Literary History*, 2, 123.
- Fish, S. (1980). *Is there a text in this class?: The authority of interpretive communities*. Harvard University Press.
- Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds & Machines*, 30, 681–694. <https://doi.org/10.1007/s11023-020-09548-1>
- Haraway, D. (1985). A cyborg manifesto. *Socialist Review*, 80, 65-108.
- Hutson, J. (2024, January 30). AI and the Creative Process: Part One. *JSTOR Daily*. <https://daily.jstor.org/ai-and-the-creative-process-part-one/>
- Iser, W. (1974). *The implied reader: Patterns of communication in prose fiction from Bunyan to Beckett*. John Hopkins University Press.
- Iser, W. (1978). *The act of reading: A theory of aesthetic response*. John Hopkins University Press.
- Iser, W. (1989). *Prospecting: From reader response to literary anthropology*. The Johns Hopkins University Press.
- Leake, B. (2021). *Derrida's style: Formalism as an address to the other* (Publication No. 24028527) [Master's thesis, Georgia State University]. ScholarWorks @ Georgia State University. <https://doi.org/10.57709/24028527>

- McGuffie, K., & Newhouse, A. (2020). The radicalization risks of GPT-3 and advanced neural language models. *arXiv*. <https://doi.org/10.48550/arXiv.2009.06807>
- Rosenblatt, L. M. (1978). *The reader, the text, the poem: The transactional theory of the literary work*. Southern Illinois University Press.
- OpenAI. (2024). [Shakespearean sonnet creation]. Retrieved [April 28, 2024], from <https://chat.openai.com/share/fcb3a141-8487-42a5-a933-5a0613f7e08f>
- Pieper, K. (2020, August 13). GPT-3, explained: This new language AI is uncanny, funny — and a big deal. *Vox*. <https://www.vox.com/future-perfect/21355768/gpt-3-ai-openai-turing-test-language>
- Watt, I. P. (1957). *The rise of the novel: Studies in Defoe, Richardson and Fielding*. University of California Press.
- Wingström, R., Hautala, J., & Lundman, R. (2022). Redefining creativity in the era of AI? Perspectives of computer scientists and new media artists. *Creativity Research Journal*. <https://doi.org/10.1080/10400419.2022.2107850>

Satya Prakash Prasad is an Assistant Professor, Stage-III (English), Department of Applied Sciences & Humanities, Jamia Millia Islamia, New Delhi.

sprasad@jmi.ac.in

Abhishek Jha is a Research Scholar at the Department of Applied Sciences & Humanities, Jamia Millia Islamia, New Delhi.

abhishekjha75@gmail.com

Shafaque Zehra is a Research Scholar at the Department of Applied Sciences & Humanities, Jamia Millia Islamia, New Delhi.

zshafaque@gmail.com