

FROM TECHNOLOGICAL TOOL TO COGNITIVE RESTRUCTURING: PARADIGM REVOLUTION IN LITERARY CRITICISM OF THE AI AGE

ZHENZHAO NIE* AND JIE REN**

Abstract *This article examines the methodological transformation that AI's entry into literary criticism has set in motion. This transformation proceeds along four interrelated dimensions—technological capacity, critical method, literary ontology, and paradigm formation—that do not unfold in linear sequence but are mutually constitutive. At the technological level, the shift from machine reading to AI reading has given literary studies a new capacity for large-scale semantic analysis, though a qualitative gap persists between AI “reading” and human reading. At the methodological level, computational analysis has expanded the scale and verifiability of criticism without being equivalent to “objectivity”; its value lies in rendering the research process more transparent and reproducible. At the ontological level, generative AI poses substantive challenges to such core categories as “author,” “text,” and “literariness,” yet this disruption extends, rather than originates, the destabilizing work already undertaken by twentieth-century literary theory. Building on these three interconnected transformations, the article proposes the paradigm of computational literary criticism,” distinguishing it from Franco Moretti’s “distant reading,” Matthew Jockers’s “macroanalysis, and the broader field of digital humanities. The article argues that the core value of computational literary criticism lies not in replacing interpretation with computation, but in constructing a collaborative framework of sustained interaction between computational discovery and humanistic interpretation—a framework whose viability depends on methodological self-discipline, data ethical awareness, and an unwavering attentiveness to the humanistic core of literary inquiry.*

Keywords: artificial intelligence, computational literary criticism, human–AI collaboration, digital humanities, literary ontology, methodological transformation

1. INTRODUCTION

We are in the midst of a paradigm revolution in literary criticism driven by artificial intelligence (AI)—a transformation this article frames as a progression from technological tool

* Yunshan Workshop of Guangdong University of Foreign Studies

** School of International Studies, Zhejiang University

to cognitive restructuring. While this technological wave originated in computer science, its impact has long transcended instrumental applications, permeating the full spectrum of the humanities and forcing a fundamental reckoning with the ontological, epistemological, and methodological foundations of literary studies. Literary criticism, the core discipline that mediates textual meaning, literary value, and the humanistic essence of literature, is at the very center of this upheaval. As algorithms begin to “read” and interpret literary texts at scale, generate literary works with coherent aesthetic and narrative qualities, and even intervene in the creative process in a quasi-agentive capacity, the cognitive foundations, methodological assumptions, and ontological presuppositions that have sustained traditional literary criticism for centuries are now placed under unprecedented scrutiny.

This upheaval manifests most acutely in two intertwined tensions that cut to the core of critical practice. The first unfolds at the level of the cognitive model and methodological framework of criticism. Traditional criticism is anchored in the close reading of individual critics, who draw on personal experience, theoretical training, and aesthetic judgment to mediate between textual form and interpretive depth. Its enduring strength lies in its capacity to unpack the nuanced contextual connotations, metaphorical richness, and affective resonance of individual texts; yet it is inherently constrained by structural limitations: unavoidable interpretive subjectivity, limited analytical scale, and a lack of systematic verifiability for its conclusions. These limitations have become increasingly pronounced in the digital age, where critics are confronted with a massive, transhistorical, and cross-media corpus of literary texts that defies the scope of conventional close reading. By contrast, AI-driven computational analysis excels at processing large-scale corpora with high efficiency and statistical consistency, identifying macro-level structural patterns and long-term historical trends that are invisible to individual readers. Even so, it remains demonstrably deficient in grasping the subtle cultural memories, ethical implications, and idiosyncratic aesthetic experiences that lie at the heart of literary texts.

The second tension goes deeper, touching on the ontological source of meaning in literary criticism. Traditional critical practice has long anchored the legitimacy of interpretation in authorial intent, the fixed textual structure, or the reader’s interpretive community—all frameworks premised on human agency as the sole source of literary meaning. This foundational premise is fundamentally destabilized in the context of AI-generated literature. When algorithms produce literary texts through the probabilistic recombination of massive training corpora, and even simulate a quasi-agentive voice in the creative process, we are compelled to revisit the most fundamental ontological questions of literary studies: Who is the author of an AI-generated literary work? Where does literary meaning originate? And what constitutes the humanistic essence of literature in an age of algorithmic creation? The interplay of these two tensions makes clear that AI’s impact on literary criticism extends far beyond the provision of new technical tools. It amounts to a systemic transformation that is reshaping the very epistemology, methodology, and ontology of the field, and thus demands a comprehensive rethinking of critical paradigms.

A substantial body of pioneering scholarship has already addressed this transformation, laying the groundwork for the field of computational literary study. Franco Moretti’s

foundational formulation of “distant reading” challenged the primacy of close reading and opened up new possibilities for large-scale literary analysis (Moretti 2013a, 48–49)¹. Matthew Jockers expanded this line of inquiry with his framework of “macroanalysis,” establishing systematic protocols for computational literary interpretation (Jockers 2013, 6–9).² Ted Underwood advanced the field by applying machine learning methods to the study of long-term literary historical change, demonstrating the explanatory power of AI for literary history (Underwood 2019, 1–11).³ Andrew Piper further refined the methodological rigor of the field, conducting a systematic scrutiny of computational validity and normative standards in literary studies (Piper 2018, 3–20).⁴ In Chinese-language scholarship, Zeng Jun has examined the interpretive logic of algorithmic criticism, while Peng Qinglong has explored the radical implications of the digital humanities for the foundational premises of literary study (Zeng 2023, 125–134; Peng 2022, 27). Nevertheless, existing research has largely proceeded from specific technological applications or singular theoretical vantage points, and the task of integrating the internal logic that connects technological foundations, critical practice, literary ontology, and theoretical paradigms remains to be accomplished.

2. TECHNOLOGICAL FOUNDATIONS: FROM MACHINE READING TO AI READING

Traditional literary reading relies on human perception and aims at the comprehension and aesthetic experience of texts—a process inseparable from affective engagement and value judgment. Yet human reading is constrained by the finitude of time, attention, and cognitive capacity, while the sheer volume of literary texts—spanning eras, regions, and languages—continues to grow. Whether for the individual reader or the scholarly community, only a limited portion can ever be read; even when confronting the same text, differences in interpretive stance, cultural context, and theoretical apparatus make exhaustive understanding unattainable. When literary textual data reaches the scale of terabytes or even petabytes, the traditional approach of relying on individual human reading becomes unsustainable in terms of time, cognition, and funding—one terabyte of text is roughly equivalent to 250,000 books, far exceeding what any individual could read in a lifetime. Although auxiliary means such as bibliographic cataloguing, literary historiography, and synoptic abstracts have enhanced the efficiency of information retrieval and overview to some degree, they have not fundamentally

¹ Franco Moretti, *Distant Reading* (London and New York: Verso, 2013), 48–49. Moretti defines “distant reading” as a research method that employs computational means to process large-scale corpora, arguing that the proper object of literary-historical inquiry should be the “literary system” rather than isolated texts—a stance that requires scholars to maintain distance from individual works and attend instead to macroscopic patterns.

² Matthew Jockers, *Macroanalysis: Digital Methods and Literary History* (Urbana: University of Illinois Press, 2013), 6–9. Jockers contrasts “macroanalysis” with “microanalysis,” contending that the former harnesses computational tools to build statistical models across tens of thousands of works, thereby revealing broad evolutionary trends in literary history.

³ Ted Underwood, *Distant Horizons: Digital Evidence and Literary Change* (Chicago: University of Chicago Press, 2019), 1–11. Underwood demonstrates the explanatory potential of machine-learning models for tracking long-durée shifts in literary style, while offering a sustained epistemological reflection on the status of computational methods.

⁴ Andrew Piper, *Enumerations: Data and Literary Study* (Chicago: University of Chicago Press, 2018), 3–20. Piper seeks to integrate data analysis with questions in literary theory, exploring concrete pathways for reconstituting the explanatory power of literary criticism on the basis of statistical evidence.

resolved the challenge posed by massive textual corpora. It is precisely this bottleneck of scale that has provided the pragmatic impetus for computational technology's entry into literary research.

Against this background, AI text analysis technologies centered on natural language processing (NLP) and machine learning (ML) have opened new pathways for literary studies. These technologies possess the capacity to process corpora rapidly and at scale, enabling the annotation and modelling of vast bodies of work in short periods, and thereby facilitating topic extraction, stylistic profiling, narrative pattern recognition, and cross-textual correlation analysis. From the perspective of technological evolution, the computer's capacity for processing text has undergone a developmental trajectory from "machine reading" to "AI reading"—a distinction crucial for understanding the manner and depth with which computational technology intervenes in literary research.

Machine reading, broadly conceived, deploys computational algorithms to perform automated analysis, organization, and information extraction, achieving structured parsing and surface-level identification. This technology has traversed three stages: from rule-based symbolic systems, through probabilistic statistical models, to early deep learning methods such as recurrent neural networks (RNN) and long short-term memory networks (LSTM) (Jurafsky and Martin 2024; Manning and Schütze 1999).⁵ Overall, machine reading's core tasks have centered on text digitization, information extraction, and basic classification; its essence consists in completing the structured processing of texts through preset rules, pattern matching, or foundational algorithms. It laid the groundwork for converting textual data into computationally tractable resources, yet its "understanding" remained at the level of symbols and structures, lacking deeper semantic insight.

AI reading, by contrast, represents a significant advance. Its technological core is rooted in breakthroughs in deep learning, particularly pre-trained language models employing the Transformer architecture (such as BERT and GPT) (Vaswani et al. 2017; Devlin et al. 2019). Large language models (LLMs), through unsupervised pre-training on massive textual corpora, map texts into high-dimensional continuous vector spaces and model probability distributions, thereby automatically acquiring the statistical regularities, semantic associations, and syntactic structures of language. As a result, AI reading is capable of deeper semantic mining rather than being confined to the recognition of surface-level information. Its core capabilities encompass semantic analysis and logical inference, sentiment computation, stylistic profiling, and knowledge association (Jin 2025, 141–148).

The transition from machine reading to AI reading constitutes a gradual evolutionary process rather than a discontinuous leap. Early deep learning methods had already begun to transcend the limitations of purely rule-based and purely statistical approaches, and while the Transformer architecture achieved a marked enhancement in capability, its underlying logic remains an extension of statistical learning and pattern matching. More precisely, the evolution

⁵ For a systematic overview of the history of NLP development, see Daniel Jurafsky and James H. Martin, *Speech and Language Processing*, 3rd ed. (draft, 2024), chaps. 6–11; and Christopher D. Manning and Hinrich Schütze, *Foundations of Statistical Natural Language Processing* (Cambridge, MA: MIT Press, 1999).

of NLP technology presents itself as a gradual spectrum from rule-driven to data-driven approaches and from surface structures to deeper semantics, within which the Transformer architecture represents a critical node but not a sharp watershed.

What, then, does this evolution of reading technology signify for literary studies? A critical examination of the metaphor of AI “reading” is warranted. AI’s processing of text is fundamentally a computational operation based on statistical patterns: the model represents words as high-dimensional vectors, computes relational weights between words through attention mechanisms, and generates output based on probability distributions. This process differs qualitatively from the subjective experience, affective resonance, cultural memory, and aesthetic judgment involved in human reading (Rayner et al. 2011; Bender and Koller 2020)⁶. As John Searle’s “Chinese Room” argument reveals, formalized symbol manipulation is not equivalent to semantic understanding (Searle 1980, 417–424). Therefore, when we say that AI can “read” literary texts, what we mean is that it can perform large-scale computational analysis of texts’ statistical structures, semantic associations, and formal features—a capability that far surpasses the human individual in scale and speed, but one that is qualitatively incommensurable with the human experience of reading.

Recognising this distinction allows us to locate more precisely the epistemological value of AI technology in literary research. This value resides not in AI’s “replacement” of human reading, but in its provision of a heterogeneous cognitive pathway for literary studies—a capacity for large-scale pattern discovery that lies beyond the reach of human cognition alone. This epistemological value operates at three levels. First, AI can process quantities of text far exceeding the limits of human reading, enabling literary studies to expand from the deep analysis of limited samples to the examination of large-scale corpora—a point that Moretti had already perceived when he proposed the concept of “distant reading.” Second, AI can identify statistical regularities and structural patterns not easily discernible through human reading, such as the evolution of topic distributions across large-scale corpora, cross-textual stylistic correlations, and long-durée semantic drift of vocabulary. When Underwood employed machine learning to examine the evolution of the English novel over two centuries, he discovered gradual trends that traditional literary-historical narratives had failed to capture adequately (Underwood 2019, esp. chaps. 2–4). Third, the process of computational analysis possesses encodable and reproducible characteristics, subjecting certain aspects of literary research to more rigorous methodological scrutiny.

Of course, large language models also suffer from the problem of “hallucination” (Ji et al. 2023, 1–38)⁷ and remain deficient in their understanding of rhetorical devices requiring deep contextual knowledge, such as irony, puns, and cultural allusions. Their analytical results are highly dependent on the quality and composition of training data; if the training corpus exhibits

⁶ On the fundamental differences between human reading as a cognitive process and AI-based text processing, see Keith Rayner et al., “Eye Movements and Word Skipping during Reading: Effects of Word Length and Predictability,” *Journal of Experimental Psychology: Human Perception and Performance* 37, no. 2 (2011): 514–28; and Emily M. Bender and Alexander Koller, “Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data,” in *Proceedings of ACL* (2020), 5185–98.

⁷ On the problem of “hallucination” in large language models, see Ziwei Ji et al., “Survey of Hallucination in Natural Language Generation,” *ACM Computing Surveys* 55, no. 12 (2023): 1–38.

biases of language, period, or genre, the analytical conclusions will be correspondingly affected. Moreover, computational analysis itself is by no means a “purely objective” operation—the researcher’s subjective judgment in corpus selection, model design, parameter setting, and the interpretation of results invariably constitutes an inseparable component of computational analysis (Da 2019; Piper 2020).⁸

The technological evolution from machine reading to AI reading does not imply that AI can “replace” human reading, nor does it imply that computational analysis is equivalent to “objective” knowledge. Yet this technological advance equips literary studies with powerful new capabilities: large-scale text processing, pattern discovery, and verifiable analysis. It provides a cognitive pathway that differs in nature from human reading yet is functionally complementary to it, thereby laying the technological and epistemological groundwork for the methodological transformation of literary criticism.

3. METHODOLOGICAL TRANSFORMATION: FROM CLOSE READING TO COMPUTATIONAL ANALYSIS

Traditional literary criticism, as an interpretive practice grounded in human cognition, is centered on the critic’s generation and reconstruction of textual meaning through individualized value judgments. This process is essentially subjective and driven by individual experience: the critic receives textual symbols, processes them through cognitive encoding, decoding, and meaning-making, and ultimately produces an interpretation shaped by the critic’s knowledge structure, cultural background, aesthetic experience, and ethical stance. As reader-response theory has shown, the plurality of textual meaning is inseparable from the situated, inevitably subjective character of the interpretive act. In this paradigm, the critic’s subjectivity serves as the crucial nexus of meaning production, while the determination of literary value relies predominantly on the critic’s sensibility, judgment, and powers of expression, and thus lacks a universally recognized and reproducible set of standards.

This close reading-centered method emphasizes the deep excavation and meticulous analysis of local textual details, privileging an internalist mode of interpretation focused on language, structure, imagery, rhetoric, and other intrinsic elements. When dealing with a single canonical text or a limited collection of works, it often demonstrates formidable interpretive depth and aesthetic acuity. However, precisely because of its high dependence on limited cognitive capacity and individual experience, this paradigm’s inherent limitations become increasingly conspicuous when confronting massive textual corpora. First, it is subject to an innate deficiency of scale: researchers, constrained by time, energy, and research orientation, can perform thick description on only a small number of texts, making it difficult to encompass broader literary phenomena. Second, it cannot fully avoid subjective preferences and selection bias in practice: the choice of texts, the focal points of detailed attention, and the paths of interpretive inference are all steered by pre-understandings and scholarly positions. Third, it

⁸ On the theory-laden nature of computational methods and the role of human interpretive choices, see Nan Z. Da, “The Computational Case against Computational Literary Studies,” *Critical Inquiry* 45, no. 3 (2019): 601–39; and Andrew Piper, *Can We Be Wrong? The Problem of Textual Evidence in a Time of Data* (Cambridge: Cambridge University Press, 2020).

frequently proves inadequate when addressing macroscopic questions such as literary-historical lineages, stylistic evolution, and the collective characteristics of group authorship—adept at microscopic explication yet limited in macroscopic synthesis. Consequently, as literary studies increasingly call for systematic, holistic, and cross-temporal investigation of large-scale corpora, the methodological lacuna of the traditional close reading paradigm becomes ever more pronounced.

The intervention of artificial intelligence offers new technological pathways for addressing the limitations of traditional criticism. Drawing on technologies such as natural language processing and deep learning, AI-driven computational criticism is effecting three fundamental transformations in critical method.

First, the perceptual mode of criticism is shifting from textual analysis to the interpretation of data visualizations—what might be called “visual reading.” With the application of computational tools and large language models, the center of gravity in literary research is moving from the direct reading of texts and documents to the interpretation of charts and data. In computational critical practice, AI disaggregates and annotates the written symbols of a text into structured data, which is then transformed into visualized outputs—sentiment fluctuation curves generated by sentiment analysis, semantic network graphs produced by topic models, syntactic structure matrices rendered by stylistic analysis, and the like. This demands that the critic acquire the capacity for “reading graphs,” decoding literary meaning from visual elements such as color, shape, and layout. A network graph of character relations in *Hamlet*, for example, can graphically display the relative prominence of characters and the density of their interactions; a word-frequency heatmap can help identify shifts in narrative density and rhythmic variation of the “solitude” theme in *One Hundred Years of Solitude* (Moretti 2013b, 220). Such visual presentations not only provide new methods for understanding literary texts but also open new avenues for the in-depth exploration of themes, characters, and plots, enabling researchers to interpret charts and data with the close attention one brings to close reading itself.

Second, the core method shifts from “interpretation” to “computation.” The fundamental change brought about by AI technology lies in driving literary criticism to construct operationalizable computational models, translating the subjective interpretation of texts into the structured computation of textual data. Grounded in probability, statistics, and pattern recognition, computational analysis translates literary features formerly apprehended through intuition into quantifiable indicators, producing evidence that is, in principle, verifiable and reproducible. Specifically, the deployment of topic models such as Latent Dirichlet Allocation (LDA) enables the automated extraction of latent thematic structures and their evolutionary trajectories within texts—revealing, for instance, the multiple distributions and shifting weights of themes such as war, love, and historical transformation in *War and Peace*. Sentiment computation technology can quantify the affective states within a text, tracing the emotional fluctuations of the “To be, or not to be” soliloquy in *Hamlet*, or tracking the shifts and turning points of anxiety in *To the Lighthouse* as a time series. Stylometric, through indices such as word-frequency distribution and syntactic complexity, provides statistical evidence for authorship disputes over the final forty chapters of *Dream of the Red Chamber*. In such ways,

aesthetic experience that resists verbal articulation is mapped onto computable, cross-verifiable multidimensional data, facilitating the interplay between humanistic appreciation and data analysis.

Third, the scope of research moves from selective sampling to corpus-wide investigation. Computational analysis transcends the sample ceiling imposed by the human reading capacity inherent in traditional close reading, rendering panoramic investigation across eras, languages, and genres possible. AI reading can process massive volumes of text beyond the reach of human readers, extending the objects of literary analysis from individual works to large-scale corpora and thereby genuinely realizing “distant reading.” Researchers could, in principle, trace the cross-textual diffusion patterns of “mechanical metaphor” across tens of thousands of Victorian novels; apply prosodic modelling to the *Complete Tang Poems (Quan Tang Shi)* to capture quantitative inflection points in the evolution of poetic form from the Early to the High Tang; and employ word-vector models for semantic topology analysis of eighteenth-century sentimental novels, revealing transformations in narrative patterns that traditional close reading could scarcely detect. The migration from “sample-based interpretation” to “comprehensive analysis” thus dismantles the established presupposition that treats texts as isolated containers, situating them instead within a dynamic sociocultural network and enhancing the comparability and reproducibility of research.

The rise of computational analysis does not entail the displacement of human subjectivity; on the contrary, it gives rise to a new paradigm of “human–AI collaboration” in critical practice. Correspondingly, the role of the researcher is reshaped: from sole interpreter, the scholar becomes the formulator of questions, the director of models, and the interpreter of meaning. Through carefully designed prompt engineering, researchers guide AI to focus on specific analytical dimensions, ensuring that theoretical self-awareness pervades the entire process of data processing and result interpretation. At the operational level, an interactive cycle of “algorithm-generated hypothesis—scholar-verified interpretation” gradually takes shape. In cross-cultural studies of magical realism, for instance, an algorithm might first identify distributional patterns of surrealist elements; the scholar then draws on geopolitical and regional knowledge to interpret these patterns, ultimately constructing a more contextually grounded explanatory model.

This collaborative mechanism gives rise to the methodology of “augmented interpretation.” Through its computational scale and pattern-recognition capacity, the algorithm can detect structures and correlations that human reading alone is unlikely to discern. The Stanford Literary Lab, for example, has employed Geographic Information Systems (GIS) to construct spatial models of the nineteenth-century global fiction market, refining the resolution of literary geography to the level of cities and neighborhoods. Similarly, applying word-embedding technology to perform semantic topology analysis on the *Complete Library of the Four Treasuries (Siku Quanshu)* could reveal differences in connection density and semantic neighbourhoods of core concepts such as “ren” (benevolence) and “yi” (righteousness) across different texts. Yet algorithms yield only patterns; it is the researcher who must situate them within cultural and historical frameworks to affect the transition from data to insight.

Driven by artificial intelligence, computational criticism is propelling literary studies toward a significant methodological transformation—shifting the focus from traditional textual reading to data visualization, from subjective interpretation toward computational modelling, and from selective sampling to comprehensive corpus-level analysis. This transformation moves scholarly judgment from subjective experience toward evidence-based and reproducible modes supported by data computation, thereby opening new horizons for understanding literary phenomena.

4. ONTOLOGICAL DISRUPTION: GENERATIVE AI AND THE RECONSTITUTION OF THE LITERARY SUBJECT

The emergence of generative AI in the literary domain constitutes not merely an updating of tools or forms but an ontological disruption. It destabilizes the author-centered paradigm of creation, gives rise to a new creative “quasi-subject,” and propels literary forms toward multimodality. Consequently, the core problem domain and the object system of literary criticism undergo a chain of reorganizations.

In traditional literary thought, the meaning and value of a work are primarily determined by the author’s biography, thought, and creative intention, and criticism accordingly privileges the analysis of creative background and psychological motivation. This “author-centrism” presupposes that literary creation is the unique expression of the author’s individual life experience, emotion, and intention. However, the rapid development of generative AI is dismantling this premise. In essence, generative AI literature is algorithmic literature, and the quality of the algorithm shapes the quality of the literary output (Zeng 2024, 163). Through the algorithmic mechanisms of large language models, Generative AI is capable of producing literary texts formerly believed to be achievable only by human authors, with linguistic, narrative, and aesthetic standards that are in many instances no less accomplished. For example, the Japanese writer Rie Kudan, composing with the assistance of Chat GPT, produced *Tokyo-to Dojo-to*, which was awarded the Akutagawa Prize, Japan’s most prestigious literary award; Professor Shen Yang of Tsinghua University completed the science fiction novel *Land of Machine Memory* with the aid of an AI platform, a work that also received an important literary prize, with the majority of the judging panel failing to detect any trace of AI involvement. These instances demonstrate that even without a human author possessing lived experience and explicit intention in the traditional sense, algorithms are equally capable of generating texts endowed with literariness and garnering professional recognition.

In the context of generative AI, the classical question of “authorial intention” is significantly attenuated. This does not amount to a simple negation of authorship but rather stems from a fundamental difference in creative mechanism: generative AI takes algorithmic production as its core, with text generation relying on the probabilistic statistics, pattern recognition, and semantic recombination of massive corpora, following a de-internationalized, non-human operational logic. By contrast, human writing is governed by individual experience, emotion, and reflection, with explicit intention serving as its pivot. Consequently, the meaning of a text no longer necessarily traces back to a pre-existing and traceable human subject but instead emerges within the configuration of algorithms, corpora, and interaction. DeepMind’s

Dramatron and other generative systems, for instance, can construct multi-branching, non-linear narrative structures under prompt-driven conditions, producing narrative labyrinths reminiscent of Borges's *The Garden of Forking Paths*. Such algorithmically driven writing of combinatorial possibility manifests non-human aesthetic characteristics that traditional literature has not fully explored. Thus, a text may well originate from an algorithmic process that is open, iterative, and devoid of traditional authorial intention. This reality invites the theoretical community to re-examine Roland Barthes's thesis of "the death of the author" and to consider its new resonances in the technological age: authorial intention is dispersed and reconstituted within the relational network of algorithmic processes, corpus ecologies, and human-machine interaction.

In the literary generation process of generative AI, the human author is to a certain degree displaced or partially supplanted, and AI thereby becomes the subject or "quasi-subject" of literary production. The rise of the "quasi-subject" depends on an entirely new creative mechanism: through representational learning from massive literary corpora, AI distills linguistic regularities and methods of textual organization and thereby constructs textual patterns that exceed human intuitive reach. For literary theory, the significance of this shift lies in the move from a closed-domain, rule-driven programming paradigm to an open-domain, data-driven statistical paradigm. By mapping texts into high-dimensional continuous vector spaces for probabilistic distribution modelling, AI autonomously learns linguistic patterns and semantic relationships, thereby acquiring capacities for comprehension and generation. The GPT series of models, for instance, by absorbing vast quantities of internet text data, have acquired rich linguistic knowledge and modes of expression, capable of generating coherent texts on the basis of learned patterns.

This mechanism prompts a reconsideration of literary "originality." As Edward Young observed in the eighteenth century, the dearth of literary originality often stems from one writer's imitation of another (Young 1759, 42). Generative AI, through autonomous learning, can internalize the diverse writing traditions and expressive resources of many periods and places within its parameter space, employing algorithms to move beyond crude imitation and thereby generating texts of novelty. Its "originality" lies not in creation ex nihilo but in the recombination, at an unprecedented scale, of existing literary resources into novel textual configurations. The "Deep Cloud" system at Guangdong University of Foreign Studies, for instance, aims to train a digital author that synthesizes the stylistic features of numerous literary masters to generate new literary works.

Generative AI's status as a creative subject has not been achieved overnight but has undergone a staged evolution from auxiliary tool to creative partner to potential independent generator. In its early stages, AI primarily assumed auxiliary functions such as grammar correction and stylistic suggestion; today, AI can participate in text generation as a creative partner—GPT-4, for example, can produce complete novel chapters, with humans providing macro-level settings through prompts, forming a collaborative loop of "human sets the framework—AI fills in the content—human revises and optimizes." As model capabilities iterate and training paradigms evolve, human micro-intervention in human-machine collaboration progressively diminishes, while AI's capacity for independent generation, quality

evaluation, and self-iteration significantly strengthens. This indicates that AI is transitioning from a passive tool toward a literary generative “agent” possessing a degree of autonomy.

Generative AI’s ontological disruption of literature is equally manifest in the transformation of literary modes of existence. While literary works have historically been anchored in written text, generative AI facilitates the production of natively multimodal works that integrate text, image, audio, and video. Stable Diffusion and DALL·E 3, for instance, can generate high-quality illustrations from textual descriptions, while AI video generation technology can transform concepts into dynamic moving images. Literature is thus shifting from purely textual reading toward a composite, immersive experience that engages visual, auditory, and other sensory modalities. This multimodal transformation is not merely a formal enrichment but a reconstitution of the mechanism of meaning production. Compared to text alone, multimodal literature possesses more complex potential for meaning generation; its information is not only more abundant but can also engender tacit knowledge (Zeng 2023, 125-134). Its meaning is not the simple addition of text, image, and sound, but is constructed through the dynamic interaction, synergy, and tension among different modalities into an entirely new meaning network and aesthetic experience. Incorporating AI video generation into literary production signifies a reshaping from medium to form and expressive conventions. Literary criticism must therefore develop new methods and capacities capable of interpreting cross-modal meaning construction.

The ontological disruption that generative AI brings to the creative subject and literary forms precipitates a series of consequential shifts in literary criticism itself. First, the objects of criticism undergo expansion and displacement: traditional criticism focused on the completed, static text, whereas now the center of gravity shifts toward the tracking and reconstruction of the generative process, such as the interpretability analysis of neural networks. The object of criticism is no longer confined to the final literary product but encompasses the entire chain of text generation—the selection and structure of algorithmic models, the composition and biases of training data, the design of prompts and feedback mechanisms, and other dynamic elements. Critics must not only analyze narrative and rhetorical structures as in close reading but must also examine algorithmic structures and parameter configurations, understanding how the system “thinks” and “creates.” Second, the core questions of criticism shift accordingly. Once the author’s status as the sole source of meaning is destabilized, “what did the author intend to express?” partially yields to “how was the text (algorithmically) generated?” and “what is the algorithmic logic of the generated text?” Criticism must investigate how algorithms learn from data and reconstruct literary patterns, how they respond to different prompts to generate texts in different styles, and the similarities and differences between their outputs and human creations in narrative structure, emotional expression, and metaphorical deployment—demanding that criticism possess corresponding computational thinking and technological literacy. Ultimately, Barthes’s thesis of “the death of the author” acquires new pertinence: generative AI does not simply proclaim the author’s absence but rather decomposes and reconstitutes “the author-function” at the technological level—its prerogatives dispersed among prompt engineers, algorithm architects and trainers, data providers, and the model itself as “quasi-subject.” Meaning production thus shifts from a single source to a pluralistic collaboration, rendering the meaning of text more open, fluid, and generative. This is not the

retreat of humanistic inquiry but rather a new demand upon criticism: to construct, under the premise of acknowledging technological reality, an interpretive framework that is both complex and inclusive, capable of productive dialogue within the tension between algorithmic generation and humanistic interpretation, thereby reassessing literary value.

Through destabilizing author-centrism, intervening in creation as a “quasi-subject,” and driving literature’s multimodal transformation, generative AI fundamentally disrupts the ontological conditions of literature. It compels literary criticism to expand its territory, revise its central problematics, and, at the intersection of technology and the humanities, rethink such fundamental propositions as “what is literature for?” and “what is criticism for?”

5. PARADIGM CONSTRUCTION: TOWARD COMPUTATIONAL LITERARY CRITICISM

The changes in method and ontology discussed above converge in what this article terms “computational literary criticism”—a paradigm that reaches beyond the mere application of technology. This marks a paradigm shift at the epistemological level of literary criticism, with its core value residing in the construction of a research chain of “data-driven inquiry—algorithmic analysis—literary interpretation.” Computational literary criticism does not aim to supplant traditional criticism; rather, informed by digital humanities, it seeks to reconfigure research pathways and cognitive frameworks, moving literary studies from subjective experiential judgment toward an evidence-based paradigm supported by data and computation, one that is verifiable and reproducible.

As an emergent paradigm, computational literary criticism exhibits several core characteristics that distinguish it from traditional criticism. First, it is grounded in data-driven foundations. AI has accelerated the digitization of literary texts, transforming works from print media into computable encoded data; computational criticism accordingly relies on large-scale corpora, converting massive bodies of work into structured data amenable to statistical analysis and modelling, and shifting the sample basis of research from limited, selective texts toward near-comprehensive collections. Analyzing the word-vector evolution of “city” imagery across 500 European novels from the seventeenth through the nineteenth century, for example, could enable the quantitative tracking of macroscopic trends in how the Industrial Revolution shaped literary spatial imagination. Data-driven approaches enhance not only the verifiability and reproducibility of research but also provide literary criticism with an expanded empirical foundation.

Second, the algorithm serves as the core mediating instrument. Computational criticism takes algorithms as its pivot; its models function both as the theoretical framework for employing computational methods to analyze literature and as concrete research tools. Whether topic modelling (LDA), sentiment computation, stylometric, or word vectors and embedding models, algorithms bear the critical responsibilities of pattern discovery and hypothesis testing. AI enables critics to detect formal and thematic patterns in texts that might otherwise go unnoticed.

Third, it emphasizes the integration of quantification and qualitative interpretation. This turn does not negate humanistic interpretation but combines subjective experience with computational methods, forming a synergistic framework of human–AI collaborative evaluation. Computational analysis does not reduce literariness to data but rather constructs, through algorithms, a literary feature space that translates formerly difficult-to-articulate aesthetic sensibilities into quantifiable, comparable, and reproducible dimensions—such as word-frequency distributions, syntactic complexity, semantic coherence, and prosodic patterns—and enhances the robustness and credibility of conclusions through statistical tests and cross-validation. Quantitative analysis provides evidence and new perspectives for qualitative interpretation, while in-depth humanistic reading endows data results with cultural significance and historical context; through mutual corroboration, the two achieve the transition from “data” to “meaning.”

Fourth, it possesses an inherent interdisciplinarity. The shift from textual analysis to visual reading compels a deep integration between literature and such fields as computer science and data science, bringing new methods and tools to literary research. Computational literary criticism draws together literary theory, natural language processing, machine learning, statistics, cognitive science, and digital humanities, forming a genuinely interdisciplinary methodological framework. It asks that researchers possess both humanistic cultivation and computational literacy, as well as an understanding of scientific norms such as experimental design, reproducibility, and open data; it thereby breaks down traditional disciplinary barriers and encourages cross-boundary innovation in knowledge production.

The transformation of literary studies toward computational analysis achieves, at the methodological level, a threefold advance. First, in research scope, it transcends the sample ceiling imposed by human reading capacity, making panoramic, cross-temporal textual analysis possible. Researchers are enabled to conduct “textual dialogue” within broader historical and geographical dimensions, revealing the long-durée evolution and cross-regional correlations of literary phenomena. Second, in analytical dimensions, it employs topic modelling, sentiment analysis, stylometric, and other techniques to transform elements such as style, affect, and theme—formerly reliant primarily on intuitive judgment—into quantifiable and verifiable systems of indicators. This quantification does not replace aesthetic judgment but provides it with references amenable to cross-examination, enhancing the transparency and debatability of interpretation. Third, in the mode of knowledge production, it forms a cognitive loop of “algorithm-generated hypothesis—scholar-verified explanation”: computational models detect latent correlations and patterns concealed within texts, while researchers draw on professional expertise to decode the literary logic and cultural implications behind the data. Criticism thus relies no longer on individual experience alone but on a form of collaborative judgment shaped through sustained engagement with algorithmic processes.

A paradigm shift necessarily entails the reconstruction of the discursive system, and computational literary criticism accordingly manifests several salient new features at the discursive level. First, critical terminology is being updated. Computational concepts are increasingly adopted to describe and analyze literary phenomena: “word vectors” to gauge semantic associations among words, “topic probability distributions” to represent the intensity

and relationships of latent themes, and “sentiment polarity” to quantify the emotional tenor of texts, among others. Second, evaluative criteria evolve accordingly. Alongside traditional aesthetic value, intellectual depth, and historical significance, computational criticism introduces new evaluative dimensions: “goodness of fit” to test the consistency between theoretical models and textual data; “pattern significance” (e.g., p-values) to assess the robustness of identified regularities; “algorithmic interpretability” to ensure that conclusions are humanly comprehensible; and “transparency and reproducibility of the computational process” as necessary conditions for scholarly rigor. These criteria bring the norms of scientific research to bear on literary criticism, reinforcing the construction of evidential chains and methodological self-awareness. Finally, all of this calls for the renewal of literary theory. The emergence of AI literature and the intervention of AI tools are reshaping not only literary forms and conceptions but also driving the innovation of research methods, with computational analysis serving as a fulcrum for reconfiguring theoretical frameworks. Traditional theories of authorship, textuality, and reader response arguably require re-examination and extension in the new context of generative AI, multimodality, and algorithmic generation. Artificial intelligence poses a new challenge for hermeneutics: to move toward “algorithmic interpretation” in literary theory, supporting interpretive activity with computable models and verifiable evidence.

At the same time, the challenges and risks attending this new paradigm deserve candid acknowledgment. Foremost among these are the inherent limitations at the technological level. First, the “algorithmic black box” problem: complex deep learning models lack transparency and interpretability in their decision-making processes, making it difficult for critics to trace and confirm how the model arrived at specific conclusions. Second, the risk of “data bias”: if training corpora harbor historical and cultural biases of gender, race, or class, model outputs may amplify and entrench these biases in analysis. Third, the concern about the attenuation of humanistic values: excessive reliance on quantitative indicators may obscure the subtle emotions, historical contexts, and individual life experiences within works that resist quantification. As Peng Qinglong has observed, “In fact, big data has, to a certain extent, already subverted the ontological and subjective nature of literature, reshaping our understanding of its humanistic character” (my translation) (Peng 2022, 27). Therefore, the essential nature of the human–machine relationship must be clarified: technology empowers rather than replaces; computation is a method, not an end. Computational tools cannot substitute for the critic’s interpretive judgment. The computational capacity of AI can be converted into reliable scholarly insight only under the guidance of the critic’s theoretical self-awareness and methodological self-discipline. The intervention of algorithms does not necessarily diminish literariness but may, on the contrary, help reveal the aesthetic potential of texts; however, the ultimate determination of value and meaning still requires human participation. Only thus can the mechanism of human–AI collaboration function effectively, ensuring that technological advancement and humanistic concern complement each other in critical practice.

6. CONCLUSION

The paradigm shift in literary criticism prompted by artificial intelligence is not simply a technological upgrade but a systemic reconfiguration of critical methods, objects, and theoretical assumptions. AI's capacity for processing massive volumes of text and parsing complex patterns provides literary studies with new instrumental resources; this drives the transformation of critical method from close reading dependent on subjective intuition and limited samples to data-driven, algorithmically mediated computational analysis, achieving the shift from sampling to comprehensive coverage, from textual analysis to visual reading, and from qualitative interpretation to the integration of quantitative and qualitative approaches. The rise of generative AI further destabilizes author-centrism, gives rise to "quasi-subject" creation and multimodal texts, and compels criticism to re-examine such ontological questions as "what is literature?" and "what constitutes authorship?" Ultimately, these transformations converge in the new paradigm of "computational literary criticism," which takes verifiable data and algorithms as its methodological foundation and, through interdisciplinary frameworks, moves literary studies toward greater empirical rigor and methodological robustness.

This paradigm shift is visible across multiple dimensions: in cognitive foundations, from reliance on the critic's subjective experience to support by data and algorithmic models; in research scale, from a focus on microscopic case studies to an expansion toward macroscopic investigation; in methodology, from traditional qualitative interpretation to "augmented interpretation" that integrates quantitative and qualitative approaches; and in disciplinary character, from relatively enclosed literary studies to convergence with computer science, data science, and other fields. It must be noted that computational literary criticism is not the terminus of critical exploration but a dynamically evolving starting point. On the one hand, it opens significant new opportunities; on the other, it brings attendant risks of algorithmic opacity, data bias, and the potential marginalization of humanistic values. Its viability lies not in the replacement of the humanities by technology but in the cultivation of a more robust relationship: computation empowers discovery, humanistic judgment calibrates meaning, and scholarly innovation advances under the twin demands of methodological rigor and sustained commitment to the values of humanistic inquiry.

Future literary scholars will arguably need to assume a dual role: on the one hand, employing computational tools and algorithmic models to discover literary patterns from massive datasets that intuition alone could scarcely grasp; on the other, maintaining the critical, historical, and value-oriented commitments of humanistic research, endowing algorithmically revealed patterns with cultural context and meaning, and exercising reflexivity toward technological processes and vigilance toward data ethics. The aspiration, ultimately, is to let computational power sharpen interpretive inquiry and to let humanistic judgment discipline the use of data—not to fuse the two into a single system but to hold them in productive, self-aware tension.

FUNDING

This work is sponsored by the project of the National Social Science Fund of China, “Research on the Construction of an Interdisciplinary Discourse System for Literary Ethical Criticism” (Project No. 19ZDA292).

ABOUT THE AUTHORS

Zhenzhao NIE is Chief Expert in “Yunshan Workshop” of Guangdong University of Foreign Studies (Guangzhou 510420, China), and an International Fellow of the British Academy and Foreign Member of the Academia Europaea. He is currently President of the International Association for Ethical Literary Criticism, specializing in Ethical Literary Criticism and interdisciplinary studies of literature (Email: niezhenzhao@163.com).

Jie REN is Associate Research Fellow at the School of International Studies, Zhejiang University (Hangzhou 310058, China). Her research focuses on Ethical Literary Criticism, modern Japanese literature, and East Asian society and civilization (Email: renjie_85@163.com).

REFERENCES

- Bender, Emily M., and Alexander Koller. 2020. “Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data.” In *Proceedings of ACL*, 5185–5198.
- Da, Nan Z. 2019. “The Computational Case against Computational Literary Studies.” *Critical Inquiry* 45 (3): 601–639.
- Devlin, Jacob, et al. 2019. “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.” In *Proceedings of NAACL-HLT*, 4171–4186.
- Ji, Ziwei, Nayeon, et al. 2023. “Survey of Hallucination in Natural Language Generation.” *ACM Computing Surveys* 55 (12): 1–38.
- Jin, Zhuo. 2025. “Literary Interpretation in Digital Humanities: Affective Computing, Digital Trust, and Their Basic Implementation.” *Journal of Social Sciences (Hunan Normal University)*, no. 2, 141–48.
- Jockers, Matthew. 2013. *Macroanalysis: Digital Methods and Literary History*. Urbana: University of Illinois Press.
- Jurafsky, Daniel, and James H. Martin. 2024. *Speech and Language Processing*. 3rd ed. Draft.
- Manning, Christopher et al. 1999. *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press.

- Moretti, Franco. 2013a. *Distant Reading*. London and New York: Verso.
- . 2013b. “Network Theory, Plot Analysis.” In *Distant Reading*, 211–40. New York, London: Verso.
- Peng, Qinglong. 2022. “Examining the Scientific and Humanistic Dimensions of Literature in the Digital Age.” *Exploration and Free Views*, no. 9, 25-28.
- Piper, Andrew. 2018. *Enumerations: Data and Literary Study*. Chicago: University of Chicago Press.
- . 2020. *Can We Be Wrong? The Problem of Textual Evidence in a Time of Data*. Cambridge: Cambridge University Press.
- Rayner, Keith, et al. 2011. “Eye Movements and Word Skipping during Reading: Effects of Word Length and Predictability.” *Journal of Experimental Psychology: Human Perception and Performance* 37 (2): 514–28.
- Searle, John. 1980. “Minds, Brains, and Programs.” *Behavioral and Brain Sciences* 3 (3): 417–24.
- Underwood, Ted. 2019. *Distant Horizons: Digital Evidence and Literary Change*. Chicago: University of Chicago Press.
- Vaswani, Ashish, et al. 2017. “Attention Is All You Need.” In *Advances in Neural Information Processing Systems* (30), 5998–6008.
- Young, Edward. 1759. *Conjectures on Original Composition*. London: A. Millar and R. and J. Dodsley.
- Zeng, Jun. 2023. “Algorithmic Interpretation: Issues in Literary Theory in the Age of Artificial Intelligence.” *Journal of Central China Normal University (Humanities and Social Sciences)* (5): 125–34.
- Zeng, Wei. 2024. “The Criticism of AI Literature: Pathways, Paradigms, and Key Issues.” *Yunnan Social Sciences*, no. 5, 163–72.