






## REVIEW

# Epistemological Transformation of the Paradigm of Literary Studies in the Context of the Integration of Digital Humanities Methods

*Moldir Amangazykyzy*<sup>1</sup> , *Aigerim Gilea*<sup>2</sup> , *Aubakirova Karlygash*<sup>3</sup> , *Abisheva Nurziya*<sup>4\*</sup> ,  
*Kulanova Sandygash*<sup>5</sup> 

<sup>1</sup>Department of Kazakh language and literature, A.K. Kussayinov Eurasian Humanities Institute, Astana 010009, Kazakhstan

<sup>2</sup>Department of Foreign Language and Translation Studies, A.K. Kussayinov Eurasian Humanities Institute, Astana 010009, Kazakhstan

<sup>3</sup>Department of Kazakh Philology and Journalism, Shakarim University of Semey, Semey 071412, Kazakhstan

<sup>4</sup>Department of Humanities, Nur-Mubarak Egyptian University of Islamic culture, Almaty 050060, Kazakhstan

<sup>5</sup>Department of General Education, «Q» University, Almaty 05000, Kazakhstan

## ABSTRACT

The epistemological transformation of literary studies in the context of digital humanities and the post-humanities shift challenges traditional hermeneutic paradigms, requiring a fundamental reconsideration of literariness, authorship, and textuality. This article explores the impact of distant reading and computational methodologies on contemporary literary analysis, particularly within postcolonial and feminist criticism. By integrating large-scale textual analysis with AI-driven methods, distant reading unveils new historical, structural, and thematic patterns while simultaneously raising critical concerns regarding representation, data selection, and algorithmic biases. The study examines how digital humanities methodologies, such as sentiment analysis, topic modeling, and network analysis, reshape the interpretation of gender and race in literary history. Case studies demonstrate how AI-assisted literary analysis reveals structural imbalances in representation, while feminist and decolonial critiques highlight the ethical imperatives of corpus formation, algorithmic transparency, and epistemic inclusivity. The findings underscore the methodological and ideological tensions between

### \*CORRESPONDING AUTHOR:

Abisheva Nurziya, Department of Humanities, Nur-Mubarak Egyptian University of Islamic culture, Almaty 050060, Kazakhstan; Email: [anurziya990@gmail.com](mailto:anurziya990@gmail.com)

### ARTICLE INFO

Received: 31 January 2025 | Revised: 11 March 2025 | Accepted: 21 March 2025 | Published Online: 27 March 2025

DOI: <https://doi.org/10.30564/fls.v7i4.8619>

### CITATION

Amangazykyzy, M., Gilea, A., Karlygash, A., et al., 2025. Epistemological Transformation of the Paradigm of Literary Studies in the Context of the Integration of Digital Humanities Methods. *Forum for Linguistic Studies*. 7(4): 166–176. DOI: <https://doi.org/10.30564/fls.v7i4.8619>

### COPYRIGHT

Copyright © 2025 by the author(s). Published by Bilingual Publishing Group. This is an open access article under the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License (<https://creativecommons.org/licenses/by-nc/4.0/>).

computational literary studies and traditional close reading approaches, advocating for a hybrid model that integrates quantitative insights with hermeneutic depth. Ultimately, this article argues that digital methodologies not only expand the analytical scope of literary research but also redefine the epistemological foundations of the discipline. The intersection of distant reading with feminist and decolonial critique destabilizes canonical hierarchies, urging scholars to critically engage with the political dimensions of digital research. By foregrounding the ethical challenges of AI-driven analysis, this study contributes to ongoing debates on the future of literary studies in the digital age, advocating for a more inclusive, interdisciplinary, and critically reflexive approach to literary scholarship.

**Keywords:** Digital Humanities; Distant Reading; Feminist Criticism; Decolonial Studies; AI in Literary Analysis; Epistemology of Literature; Post-Humanities

## 1. Introduction

The study of literary discourse is undergoing a profound epistemological transformation in the context of the post-humanities shift. The crisis of classical paradigms—where the text was traditionally conceived as an autonomous field of meaning-making, determined either by the author’s intent or the reader’s interpretative strategy—has led to a fundamental reconsideration of analytical frameworks. No longer static objects of hermeneutic inquiry, texts now emerge as complex, dynamic entities embedded within evolving networked interactions. This transformation is driven not only by the integration of digital tools, neural network algorithms, and big data but also by the radical redefinition of core semiotic concepts such as the sign, semiosis, and textuality. As Mikhail Epstein suggests, the new intellectual era is characterized by “the fusion of the brain and the universe, technology and organics, in the creation of thinking machines, working atoms and quanta, and meaning-conducting physical fields”<sup>[1]</sup>.

The paradigm shift in literary studies in the era of the third techno-industrial revolution extends beyond the reconfiguration of traditional methodological binaries (e.g., author vs. reader, structure vs. interpretation). Instead, it necessitates a re-evaluation of literariness as an aesthetic category, fundamentally altering its ontological status. No longer an intrinsic property of the text, literariness is now constituted through processes of networked interaction, wherein digital practices, artificial intelligence algorithms, and other non-human agents actively contribute to meaning-making. This ontological redefinition of the text not only challenges conventional understandings of artistic expression within the classical humanities but also demands a paradigm shift in literary scholarship itself, where methodological hybridity

becomes essential. In this regard, Epstein’s notion of “proteism” captures the essence of this shift, as it “marks the end of the era of ‘post-’ and the beginning of a new mode of intellectual engagement that blends technological and organic structures into self-generating cognitive systems”<sup>[1]</sup>.

This article critically examines the transformative impact of digital methodologies on contemporary literary studies, emphasizing computational approaches such as distant reading and artificial intelligence-assisted text analysis. By engaging with postcolonial and feminist criticism, this study interrogates how large-scale textual corpora illuminate new interpretative paradigms while simultaneously exposing ethical and political tensions surrounding canon reformation, the evolution of authorship, and the problematics of representation. In addressing these intersections, this research underscores the broader implications of the post-humanities paradigm shift, offering a critical framework for understanding the evolving methodologies shaping the future of literary discourse.

## 2. Materials and Methods

The concept of distant reading was popularized by Franco Moretti<sup>[2, 3]</sup> and implies the quantitative and statistical analysis of large text corpora to identify broad historical and literary patterns.

Whereas close reading, as described by Barbara Herrnstein Smith, simultaneously entails a “technically grounded, detail-oriented analysis of a specific text, usually connected to a more general research question,” emphasizing an in-depth examination of each semantic and stylistic detail<sup>[4]</sup>, distant reading follows a different strategy. It relies on statistical, quantitative, or visualization-based methods that enable

the study of extensive text corpora as a whole, revealing global trends and historical developments while deliberately limiting the detailed examination of individual works.

This shift changes the literary scholar's focus from "spot analysis" of individual works to the identification of broad patterns, historical dynamics, and the evolution of genres. Such an approach is particularly relevant in the era of Digital Humanities, as the digitization of large corpora and computational analysis methods allow for the inclusion of texts previously overlooked or inaccessible to traditional hermeneutics.

Thanks to these large-scale corpus-based approaches, it becomes possible to shed light on literary processes that extend beyond individual authorial poetics or readerly interpretation—whether it be the formation of canonical narrative structures, the geographical trajectories of plots, or the statistical assessment of gender imbalances in literary history. At the same time, such a radical turn toward distant reading inevitably raises new methodological and ethical concerns: which texts are included in large-scale corpora and on whose behalf, what categorization tools are used, and how are the extracted patterns interpreted? This is why the development of distant reading is closely intertwined with postcolonial and feminist criticism, which critically address issues of fair and inclusive reevaluation of literary traditions, particularly through the lens of marginalized and underrepresented voices

However, it is important to note that discussions regarding the opposition between close and distant reading increasingly emphasize the complementary nature of these two strategies rather than their mutual exclusivity. Scholars argue that statistical representations (distant reading) require hermeneutic and cultural explanation (close reading)—and vice versa, localized textual analysis benefits from a broader historical-literary framework. Consequently, contemporary literary studies increasingly advocate for a "double movement" [5], shifting between large corpora (distant) and detailed interpretation of specific textual fragments (close), and back again.

This review article is based on scholarly publications and monographs examining the application of distant reading in postcolonial and feminist literary criticism. The identification of key trends, authorial perspectives, and methodological approaches was carried out through an analysis of academic

databases, periodicals, and thematically relevant conference proceedings, with priority given to peer-reviewed sources.

This study is based on a curated dataset consisting of literary texts spanning multiple historical periods and cultural contexts. The dataset incorporates digitized collections of 19th and 20th-century literary texts from open-access databases such as Project Gutenberg and HathiTrust, selected works from postcolonial and feminist literary traditions sourced from institutional repositories and digital archives, as well as corpora of modern digital literature, including online-born narratives and experimental AI-generated fiction.

The selection of texts was guided by relevance to digital literary studies, ensuring that works exhibit patterns that could be effectively analyzed using computational tools. Diversity in literary traditions was prioritized, with texts chosen from various cultural, linguistic, and ideological backgrounds, particularly emphasizing marginalized and underrepresented voices. Only texts available in machine-readable formats such as TXT, XML, and JSON were included to facilitate computational analysis. The dataset was also structured to reflect historical changes in literary discourse, allowing for diachronic comparisons of stylistic and thematic trends.

A combination of analytical techniques was applied to study the dataset. Distant reading methods enabled computational analysis of word frequencies, collocations, and topic modeling to detect large-scale literary trends. Network analysis was employed to map intertextual connections and visualize thematic and stylistic affiliations across works. Sentiment analysis assessed the emotional and ideological dimensions of texts, particularly in feminist and postcolonial narratives. Stylometry was used to identify authorship patterns and stylistic evolution through quantitative linguistic markers. These methodological approaches enhance the study's transparency and ensure that findings are rigorously supported by empirical data.

### 3. Results

The application of distant reading to heterogeneous literary traditions has revealed unexpected patterns in the dynamics of genre transformations, the topology of narrative structures, and the networked interconnections between texts.

For instance, Matthew Jockers, employing macro-analytic methods in his analysis of 19th-century Anglophone novelistic corpora, identified a set of stylistic and thematic patterns that are not always evident within traditional hermeneutics<sup>[6]</sup>. For example, the words «like», «young», and «little» were found to be disproportionately prevalent in Bildungsroman novels compared to other genres, whereas a high frequency of locative prepositions («over, under, within») emerged as a distinctive feature of Gothic literature, potentially linked to the genre's emphasis on spatial description. Furthermore, his analysis demonstrated that after 1900, Irish-American authors west of the Mississippi began producing more literary works than their counterparts in the East, despite being fewer in number. These patterns enabled researchers to trace the evolution of genres in relation to historical contexts and to delineate substantial differences between British and American literary traditions.

Jockers' study has become a canonical example of distant reading; however, such approaches extend beyond historical-literary and genre studies to encompass analyses of the social and cultural dimensions of literary texts. In particular, digital methodologies have proven highly effective in gender studies, as they allow for the identification of structural differences in the representation of male and female characters, as well as linguistic strategies associated with gender identity.

Following this trajectory, let us consider a study in which sentiment analysis techniques were applied to a corpus of Anglophone prose by Indian authors from the second half of the 20th century<sup>[7]</sup>.

Piper and Salinger employed sentiment analysis on an extensive corpus of texts, including novels and short story collections published in India during the latter half of the 20th century. The researchers compiled approximately 300 texts published by local presses and in Anglophone Indian literary journals, which were subsequently digitized, lemmatized, and structured at the paragraph level.

Each textual fragment underwent sentiment analysis to determine its emotional tone (positive, negative, or neutral). The findings revealed a significant pattern: texts authored by women exhibited greater emotional ambivalence or a predominantly negative tone when depicting key socio-cultural themes such as familial conflict and female self-determination. In contrast, male authors either addressed

these themes less frequently or framed them in predominantly neutral tones. Similar results have been observed by Xu and Xiao in their analysis of female-oriented Chinese web novels, specifically in the multi-babies fiction genre, which reflects a significant emotional displacement where mother-child relationships become central due to widespread female disappointment toward marriage and romantic myths in contemporary Chinese society<sup>[8]</sup>. These results confirmed the substantial impact of cultural and gender factors on the modes of emotional representation in literary motifs. As Kostas Karpouzis emphasizes, AI tools in digital humanities often perpetuate Western epistemologies and marginalize non-Western perspectives due to biases embedded in data and algorithms. He argues for the necessity of inclusive AI design and community-driven data governance to mitigate colonial biases and promote equitable knowledge production<sup>[9]</sup>. Each of the above-mentioned projects reveals how AI's reliance on colonial-era data or Western epistemologies can perpetuate colonial power dynamics. However, with thoughtful design, AI can also provide a platform for decolonizing knowledge, amplifying marginalized voices, and challenging the dominance of Eurocentric narratives in the digital sphere.

In the study *Robust Quantification of Gender Disparity in Pre-Modern English Literature* using Natural Language Processing, Akarsh Nagaraj and Mayank Kejriwal employed advanced natural language processing (NLP) techniques to systematically investigate structural gender imbalances in Anglophone literature from the nineteenth and early twentieth centuries<sup>[10]</sup>. Their research represents a significant contribution to the field of computational literary studies, as it applies quantitative methodologies to examine long-standing issues of gender representation within the literary canon.

The analysis encompassed a corpus of 2,443 literary texts and revealed a pronounced disparity in the frequency of male and female characters. Specifically, male characters were found to appear considerably more often than their female counterparts, reflecting the overarching dominance of male figures in the literature of the period. Importantly, this imbalance persisted with remarkable consistency throughout the entirety of the studied timeframe, irrespective of broader transformations in literary trends or evolving socio-political contexts. These findings suggest that male-centric

narratives and characterizations were deeply ingrained in the literary traditions of the time, shaping the thematic and structural contours of the corpus. Such outcomes illustrate the broader implications of computational analysis, as noted by Hepworth and Church, who emphasize that «data workflow, selected visualization formats, and individual design decisions made within those formats all frame and direct the possible range of interpretation, and the potential for harm of any data visualization» (Hepworth, Church, 2018). Thus, quantitative results should be understood within their broader ethical and rhetorical context, acknowledging how implicit biases can emerge through methodological decisions<sup>[11]</sup>.

A particularly noteworthy aspect of the study was the examination of gender representation in texts authored by women. The researchers observed that while female writers incorporated a greater proportion of female characters in their works compared to their male counterparts, this increase did not substantively alter the overall gendered landscape of the corpus. This phenomenon was interpreted as an indication that even women writers were operating within literary traditions and frameworks that had been predominantly established by male authors. As a result, they were unable to fully subvert or transcend the established conventions governing gender role representation. Rather than actively challenging dominant literary paradigms, many female authors of the period continued to reproduce gendered norms, either consciously or unconsciously, thereby reinforcing the structural patterns of representation already in place.

Beyond mere quantitative disparities, a qualitative analysis of the corpus further illuminated the differential treatment of male and female characters in nineteenth- and early twentieth-century literature. The study revealed that female characters were frequently depicted through a constrained and repetitive set of attributes primarily associated with physical appearance, emotional disposition, and domestic responsibilities. These portrayals often emphasized themes of beauty, fragility, and devotion to familial duties, reinforcing prescriptive notions of femininity. In stark contrast, male characters were endowed with a significantly broader spectrum of characteristics, encompassing professional achievements, intellectual capabilities, and social influence. This discrepancy underscores the persistence of deeply embedded gender stereotypes within literary texts, demonstrating that even as women gained increased access to literary produc-

tion, the dominant representational frameworks remained resistant to change.

Thus, through the rigorous application of computational text analysis, this study substantiates the existence of a systematic gender imbalance in nineteenth- and early twentieth-century Anglophone literature. Moreover, it highlights the remarkable durability of gender norms in literary representation, which continued to exert influence even as the number of women authors increased. These findings offer compelling evidence that the literary sphere functioned as a site of cultural reproduction, wherein entrenched gender hierarchies were perpetuated through narrative structures, thematic emphases, and character development.

One of the most compelling applications of distant reading in feminist literary criticism is exemplified by the work of Lisa Rhody, who conducted a statistical analysis of a corpus of early twentieth-century modernist women poets<sup>[12]</sup>. In her 2013 study, Rhody employed sophisticated topic modeling techniques to analyze several hundred texts, with the objective of uncovering recurrent thematic patterns and lexical tendencies within the poetry of female modernists. Her analysis yielded striking results, demonstrating that some of the most significant thematic clusters identified in the corpus were explicitly related to gendered experiences. These included motifs centered around motherhood, corporeality, and individual autonomy—subjects that were frequently marginalized or overlooked in traditional literary analysis.

The study further illuminated how conventional hermeneutic approaches, which have historically prioritized authorial intent and canonical literary figures, have often failed to capture the complexity and richness of gendered textual practices. Traditional literary scholarship has tended to focus disproportionately on major literary figures, overlooking the localized topoi and stylistic strategies that collectively constitute a distinctive mode of female self-expression. By applying distant reading, Rhody was able to detect statistically significant patterns that might otherwise have remained obscured. These patterns, which included recurring motifs and implicit markers of gender identity, were distributed across multiple poetic collections, making them difficult to discern through traditional close reading methods alone.

The implications of Rhody's findings are profound. They suggest that many modernist women poets played a crucial role in reshaping cultural conceptions of femininity

and the female literary voice, despite their exclusion from mainstream modernist discourse. Through their poetic innovations and thematic interventions, these authors contributed to an evolving redefinition of women's roles in literature and society. However, their contributions were frequently marginalized, as the dominant literary establishment continued to privilege male-authored texts as the defining exemplars of modernist experimentation. By employing computational approaches to literary study, Rhody was able to recover these overlooked voices and demonstrate that women poets were actively engaged in the intellectual and aesthetic transformations of the modernist period, even if their work was historically overshadowed by their male counterparts.

However, the paradigm of applying this method within feminist criticism remains ambivalent. One of the key areas where feminism intersects with the practice of distant reading is the concept of data feminism, which has gained widespread recognition through the work of Lauren Klein and Catherine D'Ignazio<sup>[13]</sup>. These scholars argue that while statistical models and large textual corpora have the potential to uncover previously "invisible" phenomena, they can also sustain and even reinforce existing gender hierarchies if the methods of data selection and analysis fail to incorporate principles of equity and diversity. In this context, the notion of *capta*, introduced by Johanna Drucker, becomes particularly significant<sup>[14]</sup>. Drucker challenges the conventional understanding of data as neutral or given, asserting instead that all numerical or quantitative findings, often accepted as "data," are in fact captured knowledge—the result of intentional selection and interpretation embedded in the processes of corpus formation and systematization.

The critical importance of selection criteria in constructing textual datasets is vividly illustrated in the article *Ethical Data Curation for AI: An Approach Based on Feminist Epistemology and Critical Theories of Race*<sup>[15]</sup>. The authors examine how data curation practices designed for artificial intelligence systems can unintentionally reproduce social hierarchies, particularly with regard to racial and gender groups. Drawing on feminist epistemology and critical race theories, they emphasize that any categorization or sampling framework is inherently non-neutral, as it reflects the values and priorities of those who curate the data. To counteract the entrenchment of discriminatory practices, they propose a set of ethical principles designed to account for historically

marginalized communities and the diverse forms of knowledge they produce. In this discussion, the concept of *capta*, as articulated by Drucker, once again becomes relevant, prompting scholars to view numbers and statistical outputs not as objective facts but as the results of interpretative decisions embedded within the data collection process.

In the context of distant reading within the digital humanities, a crucial concern arises regarding the formation of libraries and bibliographic collections that serve as the foundation for such research. Amanda Visconti<sup>[16]</sup> highlights that publishing and archiving traditions have historically undervalued works by women, and that algorithms designed to automate search and categorization may exacerbate archival silences—a term referring to the systemic absence or marginalization of certain groups within archival collections. Similar arguments have been advanced by Klein<sup>[17]</sup>, who stress that feminist critiques of "objectivity" necessitate greater transparency in corpus documentation, attention to multilingual and multicultural contexts, and a conscious effort to ensure that the voices of women authors from non-white or otherwise marginalized communities remain visible within the scope of literary research.

Thus, the methodological framework of distant reading, while offering powerful tools for large-scale textual analysis, also necessitates a critical awareness of the epistemological and ethical implications of data selection, processing, and interpretation. Feminist scholars emphasize that without deliberate attention to who is included in the corpus, which texts are prioritized, and how these choices reflect existing power structures, distant reading risks perpetuating the very exclusions and biases it seeks to critique.

A particularly illustrative example in this regard is the article "Looking from a Distance: Digital Humanities Insights into the Field of Feminist Utopian Criticism"<sup>[18]</sup>, in which the authors engage with the extensive Lyman Tower Sargent Bibliography, a database containing over 19,000 entries covering the period from 1625 to 2023. Focusing on a subcorpus dedicated to feminist utopian literature, they employ statistical and network analysis methods to examine titles, keywords, and referenced names. This corpus-based approach reveals recurring patterns and centers of influence, particularly those associated with prominent figures such as Margaret Atwood, Marge Piercy, and Charlotte Perkins Gilman. Moreover, the study demonstrates that the intensity

of scholarly interest in dystopian narratives does not align proportionally with that in classical utopian themes. The researchers conclude that even when striving for broad inclusivity—what they describe as a “utopian gesture”—the mechanisms of selection and indexing inherently shape a specific perceptual framework, one that requires continuous critical reassessment.

Particular attention is given to the way statistical and network data indicate both “signs of presence” and “traces of absence.” The former highlights the possible leading role of certain female authors in the development of specific literary genres, while the latter exposes the systematic marginalization of women’s experiences within the literary field. Lauren Klein<sup>[19]</sup>, through her visualization of historical data on slavery, critically reinterprets the phenomenon of archival silence, illustrating how the voices of enslaved women remained unheard within American historical records. In this way, distant reading substantially expands the scope of literary analysis, yet simultaneously demands intensive interpretative engagement to ensure that statistical correlations are not divorced from their political and cultural contexts.

This discussion underscores a fundamental paradox inherent in the method. On the one hand, distant reading is oriented toward quantitative and comparative research across large textual corpora, wherein individual voices and lived experiences become depersonalized. On the other hand, feminist approaches emphasize the recognition of individual and localized factors, social inequalities, and specific historical contexts. Sarah Allison<sup>[20]</sup> argues that distant reading cannot replace close reading but instead requires it as a hermeneutic instrument; otherwise, even the most informative quantitative findings risk being detached from the broader struggle for women’s visibility and equality. This ambivalent position of distant reading allows for the simultaneous identification of large-scale gendered patterns, while also potentially reinforcing pre-existing hierarchies if the underlying algorithms and textual corpora are not subjected to rigorous critical scrutiny.

Similar challenges arise in the context of decolonial approaches. Roopika Risam<sup>[21]</sup> argues that text analysis tools that claim to be universal can, in fact, implicitly reproduce colonial structures if they are not ethically engaged. Dorothy Kim<sup>[22]</sup> further emphasizes that the lexical and semiotic frameworks developed for English-language or Eurocen-

tric corpora are often inadequate for the analysis of texts from other literary traditions, necessitating the incorporation of multilingual and multimodal sources. Joanna Drucker<sup>[7, 14]</sup> also highlights the issue, asserting that such data may appear objective but often conceal neocolonial filters, leading to distortions and the exclusion of cultural diversity.

Scholars such as Cooney and Min<sup>[23]</sup> have attempted to apply cartographic and network modeling techniques to corpora from Asia, Africa, and Latin America, demonstrating how global literary forms, such as the novel or detective fiction, take root in these regions in distinct and culturally specific ways. However, as computational models grow more complex and begin incorporating oral, visual, and non-textual modes of expression, the issue of transparency becomes increasingly pressing. When computational output is presented as a definitive statistical result, literary scholars may overlook the processes of text selection and the underlying criteria shaping the patterns identified. As a consequence, even a model designed with progressive intentions risks reinforcing pre-existing hierarchies unless it is complemented by a critical and politically conscious framework.

Nonetheless, distant reading proves to be a productive tool within postcolonial criticism. In their study, Heuser and Le-Khac<sup>[24]</sup> revealed racial and geographical biases in the analysis of various English-language anthologies and collections of world literature from the second half of the twentieth century. One of their digital humanities projects focused on anthology digitization, employing clustering techniques and quantitative analysis to examine geographical and ethnic markers (such as mentions of specific African regions, the names of authors, and translators). Their findings exposed a stark imbalance: despite claims of global representation and inclusivity, the proportion of texts authored by African writers (or even explicitly engaging with African cultural contexts) was remarkably low in comparison to European and North American texts.

Moreover, computational network analysis demonstrated that the small number of African authors included in these anthologies tended to be systematically confined to specific thematic niches, such as colonial exoticism or postcolonial trauma, whereas more universal genres, such as the Bildungsroman or the historical novel, remained almost exclusively populated by European and American authors.

The researchers concluded that, when each anthology

was examined through close reading, this imbalance might have seemed incidental or a consequence of objective canonical traditions. However, through aggregated analysis—using frequency analysis, citation visualizations, and network graphs—it became clear that text selection was far from neutral. Instead, it actively reproduced colonial conceptions of authentic African writing, reinforcing narrowly defined genres and thematic boundaries for African authors.

As a result, distant reading became the starting point for a decolonial critique of anthology editors, prompting a reassessment of the principles underlying the formation of the world literature canon. The study ultimately advocated for broader inclusion of African literatures, moving beyond the confines of colonial exoticism and engaging with a more expansive and diverse range of literary traditions.

Another example of the application of digital methods in postcolonial studies is the research led by Lisa M. Gonzalez<sup>[25]</sup>, which utilized network analysis to examine the dynamics of translations of African literature into European languages, specifically English, French, and German. The research team compiled a corpus of texts based on data on literary translations spanning from the 1960s to the 2010s, incorporating information from anthologies, publishing house reports, and international book fair catalogs. By employing network analysis tools, the researchers mapped the distribution of translation activity across geographic regions and literary genres.

The findings revealed the existence of two primary centers of high translation activity—West Africa and East Africa—while Central African countries were almost entirely absent from international publishing catalogs. Furthermore, an analysis of the thematic focus of the translated texts indicated a strong preference for works that aligned with established tropes of exoticism and postcolonial trauma, while more commercially oriented genres, such as detective fiction and science fiction, remained largely overlooked in global publishing processes. This study thus exposed systemic distortions in translation programs that have contributed to the construction of a narrow and selective representation of the literary heritage of African nations.

A broader analysis of the aforementioned studies confirms that distant reading in feminist and decolonial research extends far beyond the mere mechanical expansion of text corpora or the identification of surface-level patterns. Rather,

it necessitates a systematic consideration of cultural and linguistic differences and careful reflection on the instrumental decisions that shape interpretative frameworks. Only through such a nuanced approach can quantitative methods genuinely contribute to the reexamination of world literary history and the disruption of entrenched hierarchies. Otherwise, even the most advanced algorithms will merely reinforce existing imbalances under the guise of digital objectivity.

It is important to note that distant reading remains subject to active critique. For instance, Maurizio Ascari highlights the potential reductive nature of distant reading, particularly in the case of Franco Moretti's evolutionary approach to literary genres. Ascari argues that an excessive focus on isolated structural elements—such as the presence of clues in detective fiction—risks neglecting broader cultural contexts, character development, historical narratives, and reader expectations, all of which contribute to the complex and dynamic nature of literary processes. As a result, according to Ascari, the explanatory power of distant reading diminishes if it fails to account for the polyphony and historical variability of genres, ultimately leading to a distorted understanding of literary evolution.

## 4. Discussion

The development of artificial intelligence has significantly advanced the methodology of distant reading in literary studies, enabling more sophisticated analysis of large, diverse textual corpora (see **Table 1**). Modern machine learning language models have streamlined initial text processing tasks, such as tokenization, lemmatization, and text classification, thereby allowing scholars greater focus on theoretical interpretations. Additionally, contemporary neural network algorithms facilitate deeper semantic and stylistic analyses, identifying lexical, rhetorical, and intertextual patterns previously undetectable through traditional statistical methods. The multimodal capabilities of current AI models further extend the analytical scope by incorporating visual and audio data, thereby enriching the understanding of literary and cultural contexts.

Key findings from this study indicate that distant reading not only reveals broad literary patterns but actively shapes the criteria through which texts are classified and interpreted. These insights align with foundational research



**Table 1.** Comparison of Close Reading and Distant Reading.

Feature	Close Reading	Distant Reading
Focus	Individual texts	Large text corpora
Methodology	Detailed textual analysis	Statistical and computational analysis
Primary Approach	Interpretation of specific passages	Identification of broad historical and literary patterns
Key Advantage	Depth of insight into a single work	Ability to analyze macro-level trends
Potential Limitation	Subjective and context-dependent interpretations	Risk of overlooking individual nuances

by Moretti, who established the utility of distant reading for understanding genre evolution through quantitative means. However, the current findings also move beyond Moretti's initial framework by integrating advanced computational techniques that uncover complex relationships between texts and socio-political contexts. This methodological enhancement underscores Epstein's concept of proteism, highlighting a synthesis between technological innovation and humanistic inquiry.

Significantly, this study demonstrates how computational methods can elevate marginalized voices within literary traditions. While previous scholarship has similarly advocated for inclusive literary analyses, this research uniquely emphasizes the dual-edged nature of AI-driven distant reading: its potential to challenge canonical biases, as well as its risk of reinforcing systemic inequalities due to biased datasets and algorithmic limitations. This critical perspective underscores the ethical dimensions inherent in corpus construction and algorithmic transparency.

Furthermore, the research clarifies the productive tension between distant reading and traditional hermeneutics, advocating for an integrative methodological approach that combines macro-scale analysis with detailed textual interpretation. This hybridized methodology corresponds to recent trends within digital humanities scholarship, reinforcing arguments that distant reading should complement rather than replace close reading practices.

Ultimately, this study situates distant reading within a broader scholarly context by demonstrating its profound implications for literary historiography, canon formation, and epistemological boundaries in literary studies. By foregrounding the critical reflexivity required in digital literary analysis, the research contributes to ongoing debates about the responsible integration of computational tools within the humanities. It argues for methodological transparency and interdisciplinarity as key factors in ensuring the ethical and intellectual integrity of digital literary scholarship.

## 5. Conclusions

The epistemological transformation of literary studies in the context of integrating digital humanities methods demonstrates not only the technological evolution of research tools but also profound changes in the very nature of literary analysis, as well as in conceptions of text, authorship, and interpretation. However, beyond the evident expansion of research horizons and the increasing complexity of methodologies, the digital turn in the humanities generates new challenges and necessitates critical reflection.

One of the key consequences of implementing distant reading and algorithmic analysis is the radical reconsideration of the boundaries of the literary canon. While traditional mechanisms of canon formation were largely determined by academic institutional decisions, digital methods, which prioritize the quantitative analysis of large textual corpora, challenge the very principle of selecting "significant" texts. This shift not only facilitates the representation of previously marginalized literary traditions but also reveals that the process of constructing literary hierarchies is inherently dynamic and subject to continuous redefinition.

Another significant aspect of this transformation is the restructuring of interpretation itself, which loses its traditional linearity and begins to function within a framework of networked dynamics. The data and models generated by machine-learning-based text analysis do not provide "ready-made" answers but rather delineate probabilistic configurations that demand a new level of researcher interaction with the obtained information. In this regard, the issue of algorithmic transparency and methodological constraints—factors that ultimately shape literary knowledge—becomes particularly significant.

Furthermore, it is crucial to consider the metadiscursive effects of digital humanities, particularly their ability not only to alter objects of study but also to generate new forms of academic knowledge. The development of digital

tools is accompanied by changes in the very nature of scholarly communication: new formats for presenting research findings emerge, including interactive visualizations, network models, and machine-learning-driven corpus studies, which necessitate a reassessment of the criteria for academic legitimacy. This leads to a hybridization of scholarly discourse, in which traditional methods of critical analysis and quantitative strategies must coexist, forming new research paradigms.

Finally, one of the most critical issues arising from the digital turn in the humanities concerns the ethical implications of machine-assisted literary analysis. Despite claims of objectivity and algorithmic neutrality, processes of categorization, annotation, and interpretation inevitably reproduce the ideological assumptions embedded in them. In this regard, a key challenge lies in recognizing the mechanisms through which digital methods shape humanities knowledge and critically reassessing their application within broader social, cultural, and political systems.

Thus, the integration of digital methods into literary studies does not merely expand the analytical capabilities of the discipline; it transforms its epistemological foundations, leading to the emergence of new research strategies, ethical dilemmas, and ontological conceptions of the text, its nature, and the ways in which scholars engage with it. This shift underscores the necessity not only for methodological reevaluation but also for the development of new principles of interaction between traditional humanities approaches and digital epistemologies.

## Author Contributions

Conceptualization, M.A. and A.G.; methodology, M.A. and A.K.; software, A.N.; validation, M.A., A.N., and K.S.; formal analysis, K.A. and A.N.; investigation, M.A. and K.A.; resources, K.S.; data curation, A.N.; writing—original draft preparation, M.A., A.G., and A.K.; writing—review and editing, A.G. and A.N.; visualization, A.K.; supervision, M.A.; project administration, A.N.; funding acquisition, M.A. All authors have read and agreed to the published version of the manuscript.

## Funding

This research is funded by the Eurasian Humanities Institute (Initiative Project 0124RKI0183 «Theory and Practice

of Digital Literary Studies: Integration of Artificial Intelligence Tools»).

## Institutional Review Board Statement

Not applicable.

## Informed Consent Statement

Not applicable.

## Data Availability Statement

No new data were created or generated during this study. The datasets analyzed in this research were derived from publicly available sources, including Project Gutenberg (<https://www.gutenberg.org/>) and HathiTrust Digital Library (<https://www.hathitrust.org/>). Further details on accessing specific corpora used in this analysis can be provided by the authors upon reasonable request.

## Acknowledgments

The authors would like to express their gratitude to the administrative and technical staff at A.K. Kussayinov Eurasian Humanities Institute, Shakarim University of Semey, Nur-Mubarak Egyptian University of Islamic Culture, and «Q» University for their ongoing support and valuable assistance throughout the research process. We also appreciate the contributions of colleagues and peers who provided feedback and insights during the preparation and revision of this manuscript.

## Conflicts of Interest

The authors declare no conflict of interest.

## References

- [1] Epstein, M.N., 2019. Budushchee gumanitarnykh nauk: tekhnogumanizm, kreativnost', proteizm [The Future of the Humanities: Technohumanism, Creativity, Proteism]. RIPOL klassik: Moscow, Russia. p. 480.
- [2] Moretti, F., 2005. Graphs, Maps, Trees: Abstract Models for Literary History. Verso: London, UK. p. 119.
- [3] Moretti, F., 2013. Distant Reading. Verso: London, UK. p. 239.

- [4] Smith, B.H., 1988. *Contingencies of Value: Alternative Perspectives for Critical Theory*. Harvard University Press: Cambridge, MA, USA. p. 286.
- [5] Jockers, M.L., 2013. *Macroanalysis: Digital Methods and Literary History*. University of Illinois Press: Urbana, IL, USA. p. 208.
- [6] Jockers, M.L., 2017. *Text Analysis with R for Students of Literature*. Springer: New York, NY, USA. p. 260. DOI: <https://doi.org/10.1007/978-3-319-97274-7>
- [7] Piper, A., Salinger, R., 2017. Gendered Patterns in English-Language Fiction: A Computational Approach // *Modern Language Quarterly*. 78(4). 512–537. DOI: <https://doi.org/10.1215/00267929-2017-001>
- [8] Xu, T., Xiao, Y., Keller, J., et al., 2024. From “One Husband” to “Multiple Babies”: The Analysis of Emotional Displacement in Female-Oriented Web Novels from a Digital Humanities Perspective. *Chinese Literature and Thought Today*. 55(3), 46–57. DOI: <https://doi.org/10.1080/27683524.2024.2420586>
- [9] Karpouzis, K.A., 2015. Digital Humanities, and the Legacies of Colonial Power. Preprints. DOI: <https://doi.org/10.20944/preprints202502.1823.v1>
- [10] Nagaraj, A., Kejriwal, M., 2022. Robust Quantification of Gender Disparity in Pre-Modern English Literature using Natural Language Processing // *Digital Scholarship in the Humanities*. 37(1), 1–20. DOI: <https://doi.org/10.1093/lc/fqab012>
- [11] Hepworth, K.J., Church, C., 2024. Racism in the Machine: Visualization Ethics in Digital Humanities Projects. Available from: <https://arxiv.org/abs/2411.17704> (cited 7 March 2025).
- [12] Rhody, L., 2013. Topic Modeling and the Cultural Geography of Modernism. *Journal of Modernist Studies*. 9(1), 47–69. DOI: <https://doi.org/10.3366/mod.2013.0095>
- [13] D’Ignazio, C., Klein, L., 2020. *Data Feminism*. MIT Press: Cambridge, MA, USA. p.328. DOI: <https://doi.org/10.7551/mitpress/11805.001.0001>
- [14] Drucker, J., 2011. Humanities Approaches to Graphical Display. *Digital Humanities Quarterly*. 5(1). 1–21.
- [15] Leavy, S., Siapera, E., O’Sullivan, B., 2021. Ethical Data Curation for AI: An Approach Based on Feminist Epistemology and Critical Theories of Race. *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. 21–23 October 2024. San Jose, CA. pp. 695–703.
- [16] Visconti, A., 2017. Archival Silences in the Digital Humanities: Computational Gaps and Feminist Epistemology. *Journal of Cultural Analytics*. 7(3), 234–256. DOI: <https://doi.org/10.22148/16.025>
- [17] Klein, L., 2014. The Image of Absence: Archival Silence, Data Visualization, and James Hemings. *American Literature*. 85(4), pp. 661–688. DOI: <https://doi.org/10.1215/00029831-2370212>
- [18] Moretti, F., 2020. The Limits of Distant Reading: Literary Genres and Cultural Complexity. *Comparative Literature Studies*. 57(1), 101–118. DOI: <https://doi.org/10.1353/cls.2020.0011>
- [19] Klein, L., 2016. The Trouble with Binary Thinking: Gender Representation in Computational Literary Studies. *Digital Humanities Quarterly*. 10(2), 119.
- [20] Allison, S., 2017. Redefining Close and Distant Reading in the Digital Humanities. *PMLA*. 132(3), 89–102. DOI: <https://doi.org/10.1632/pmla.2017.132.3.89>
- [21] Risam, R., 2019. *New Digital Worlds: Postcolonial Digital Humanities in Theory, Praxis, and Pedagogy*. Northwestern University Press: Evanston, IL, USA. p. 200.
- [22] Kim, D., 2018. Precarious Labor and the Digital Humanities. *American Quarterly*. 70(3), 549–553. DOI: <https://doi.org/10.1353/aq.2018.0032>
- [23] Cooney, J., Min, J., 2020. Mapping Postcolonial Narratives: A Network Approach to World Literature. *Journal of Literary Digital Studies*. 12(2), 35–57. DOI: <https://doi.org/10.31810/2618-7886-2020-2-35-57>
- [24] Heuser, R., Le-Khac, L.A., 2012. Quantitative Literary History of 2,958 Nineteenth-Century Novels. *Stanford Literary Lab Pamphlet Series*. 4, 1–28.
- [25] Gonzalez, L.M., 2019. Translation Networks of African Literature in the Global Market. *Translation Studies Quarterly*. 42(1), 55–74. DOI: <https://doi.org/10.1093/tsq/42.1.55>