

Artificial Intelligence in the humanities: Between innovation and challenges

Nurali Chelebiyev*¹ 

¹Azerbaijan State Pedagogical University, Sheki Branch,
Sheki, Azerbaijan

Abstract. The purpose of the study was to determine the boundaries and possibilities of using artificial intelligence technologies in the examination of literary and historical sources. The methodological strategy combined critical analysis of published findings, systematisation of tools, evaluation of representative practices in the digital humanities, and integration of these observations within a unified analytical framework that links technical procedures with interpretative conclusions. Particular attention was paid to case studies that illustrate the capacity of algorithms to reconstruct latent narratives and intertextual networks. The obtained findings demonstrate that topic modelling combined with clustering reliably identifies latent themes and their temporal dynamics in large text collections; contextual neural architectures provide consistent identification of persons, toponyms, and genre features across dispersed corpora; stylometric approaches, together with tools for detecting textual correspondences, enable localisation of shifts in authorial signal and reconstruction of borrowing chain. The study established that tools of geographical annotation allow the reconstruction of spatial trajectories of characters, correspondence, and routes. The analysis also indicated sensitivity of results to the quality of digitisation: errors in the recognition of printed and handwritten text altered thematic distributions and affected attribution accuracy, which required prior cleaning and normalisation. The study noted that generative systems are appropriately considered instruments of co-creation at three levels: operational (variation of formulations), procedural (structural organisation and iterative support), and conceptual (semantic integration and interpretative framing), while decisive responsibility remains with the human researcher. The ethical dimension confirmed the presence of risks related to bias in training collections, threats to confidentiality, and potential distortion of cultural meanings; requirements for transparency, documentation of the role of automated tools, and the involvement of domain expertise were outlined. The practical value of the study lies in the fact that the proposed framework provides researchers with a clear sequence of actions for reproducible analysis of textual corpora, ensures accurate attribution and interpretative validity, and offers instrumentally grounded guidelines for the secure integration of technologies into humanities projects

Keywords: topic modelling; intertextual connections; authorship attribution; discursive features; handwritten texts; variation of formulations

INTRODUCTION

Research on the interaction between AI (Artificial Intelligence) and the humanities acquires particular significance in connection with the expanding use of algorithms for the analysis of cultural texts, historical sources, and artistic practices. The need to examine this issue arises from the duality of technological impact: on the one hand, such systems introduce new instruments for reconstruction, attribution, and interpretation; on the other, they raise questions concerning the reliability of results, their

compliance with the methodological standards of the humanities, and the ethical implications of integrating algorithms into the research process.

The limitations and potential of algorithmic approaches were examined in detail by W. Duch (2024), who emphasised that AI technologies could substantially enhance the analytical tools of humanities researchers, yet do not fully reproduce the contextual depth and value-based multidimensionality of humanistic knowledge. The researcher

Suggested Citation:

Chelebiyev, N. (2025). Artificial Intelligence in the humanities: Between innovation and challenges. *Journal of Digital Humanities and Cultural Innovation*, 1(1), 4-15. doi: 10.67524/verus5/1.2025.04.

*Corresponding author (chelebiyevnurali@gmail.com)



noted that the effectiveness of machine methods is constrained by the absence of interpretative capacity within such systems. The academic field increasingly regards intelligent systems as a key driver of the development of digital humanities. E. Frontoni *et al.* (2024) stressed that these technologies open a new horizon for digital humanities practices, combining rapid processing of large corpora with the integration of complex models for reconstructing intertextual and spatial connections. The researchers maintained that the potential of these tools derives from their capacity for multi-level analytics, although they remain dependent on human interpretation and oversight.

An important line of inquiry concerns the transformation of authorship in the era of algorithmic systems. S. Colella (2025) observed that texts generated with the assistance of language models enter into a complex relationship with conventional conceptions of authorship, since their form and structure could reproduce features of human creativity, whereas the absence of authorial intention and responsibility places them within a grey zone of cultural practice. The researcher observed that such texts alter the understanding of the performativity of the authorial act and raise questions regarding the boundaries of authentic writing. Research on the use of algorithms in the preservation and interpretation of cultural heritage also proved significant. E.D. Hannaford *et al.* (2024) demonstrated that the application of AI tools to digital collections supports the development of infrastructures for community archives and connected local histories with broader narratives. The researchers concluded that machine learning systems effectively support classification and retrieval processes in archives, yet require community oversight to prevent distortion and preserve the authenticity of cultural data.

Several studies emphasise the need for changes in training systems for humanities research. J. Chun & K. Elkins (2023) proposed a curriculum model grounded in the principle of humanities-oriented AI, which aims not only at technical instruction in working with algorithms but also at cultivating critical thinking capable of assessing the limits and risks of such systems. The researchers concluded that researcher training requires integration of technical literacy and ethical reflection. K. Lavidas *et al.* (2024) showed that the readiness of students in the humanities and social sciences to use AI for academic purposes is shaped by a range of factors, from technical awareness to individual expectations and prior experience with digital tools. The researchers established that integration of such tools into educational processes is feasible only when educational and cultural contexts are carefully considered, which underscores the relevance of implementing these technologies in the humanities not only at the level of research but also of education.

A considerable contribution to the discussion of the dual character of intelligent systems in the digital humanities was made by H. Zhao (2023). The researcher argued that the use of algorithms generates serious

challenges while opening new possibilities: such systems could accelerate analysis and expand access to large corpora, yet risk reproducing bias and distorting outcomes. The debate also addressed the transformative potential of algorithms in education and research. A. Shahid *et al.* (2024) emphasised that AI has the capacity to reshape educational practices in the humanities by enabling personalisation, automation of routine tasks, and broader access to resources. The researchers noted that these developments exert a long-term influence on the structure of educational programmes and the organisation of scholarly knowledge.

Despite the substantial body of existing studies, gaps requiring further examination remain. The relationship between algorithmic efficiency and interpretative validity remains insufficiently clarified. The ways in which authorship and creativity are transformed under the influence of generative systems within academic and cultural practices also remain underexplored. Ethical dimensions require further specification, including risks of bias, issues of confidentiality, and challenges of cultural representation. In response to these identified issues, the purpose of the study was to evaluate the effectiveness of algorithmic approaches in the examination of literary and cultural sources and to delineate the limits of their use for formulating academically grounded conclusions. The objectives of the study were as follows: to identify which machine learning algorithms are applied most effectively to the analysis of literary and historical texts; to examine the transformation of authorship and creativity in the digital age under the influence of generative systems; and to analyse ethical challenges related to bias, confidentiality, and cultural representation.

MATERIALS AND METHODS

The study employed a comprehensive methodology combining critical analysis of publications, evaluation of digital tools, and systematisation of practices in digital humanities research. The study unfolded in several sequential stages. The first stage identified algorithmic approaches of central importance for the analysis of corpora of literary and historical texts. The next stage examined examples of practical application of these approaches in specialised projects. A separate section addressed tools for analysing spatial and intertextual dimensions of texts, along with instruments for processing manuscript and digitised sources. The subsequent stage focused on questions of authorship and creativity arising from the spread of generative systems. The theoretical framework was strengthened through engagement with the cultural concepts of J. Lotman (semiosphere) (Bankov, 2020), Z. Bauman (liquid modernity) (Doyle & Conboy, 2020), and A. Appadurai (cultural flows) (Linares, 2025), which enabled interpretation of the analytical results not merely as technical outcomes but as elements of broader cultural processes. Tools and case studies were grouped by thematic direction in tables (Table 1).

Table 1. Methods, tools, and cases of AI application in digital humanities research

Algorithms for text corpus analysis			
Method/tool	Task/function	Example of application	Source
Latent Dirichlet allocation (LDA)	Topic modelling, identification of latent topics	Classification of genre-thematic configurations in large corpora	U. Chauhan & A. Shah (2021)
Stylometry (R stylo, Δ -metrics)	Authorship attribution, style analysis	Detection of changes in authorial signal within a single work	T. Merriam (2021)
Transformers (BERT – Bidirectional Encoder Representations from Transformers; Longformer, BigBird)	Named entity recognition, classification of discursive features	Analysis of long texts (novels, periodicals)	A.H. Mohammed & A.H. Ali (2021); H. Miao <i>et al.</i> (2024)
Tools for intertextuality and spatial dimensions			
Recogito	Geographic reconstruction, working with toponyms	Mapping routes in texts	Pelagios Network (n.d.)
textreuse (R), passim	Detection of repetitions and borrowings	KITAB (Knowledge, Information Technology, and the Arabic Book), analysis of Arabic corpora	M. Romanello & S. Hengchen (2021)
Tesserae	Search for allusions and intertextual links	Latin and Ancient Greek corpora	Tesserae (n.d.)
Platforms for working with manuscripts and digital archives			
Transkribus	Handwritten text recognition (HTR – Handwritten Text Recognition)	Manuscripts of J. Bentham	Project Update... (2018)
ECCO (Eighteenth Century Collections Online)	Assessment of OCR/HTR noise (OCR – Optical Character Recognition; HTR – Handwritten Text Recognition)	Analysis of 18 th -century corpus	A. Liimatta (2024)
HathiTrust Research Center (HTRC)	Large-scale non-consumptive corpus analysis	Development of worksets from thousands of volumes	HathiTrust (n.d.); HathiTrust Research Center Analytics (n.d.)
Practical application			
Mining the Dispatch	Topic modelling of press corpus	Reconstruction of Civil War discourses in the United States	J. Longstreet (2024)
The Viral Texts Project	Detection of networked text circulation	19 th -century newspaper republications	The Viral Texts Project (n.d.)
KITAB	Detection of borrowings in classical texts	Arabic-language corpora	KITAB (n.d.)
ReRites	Co-creation experiments	Text generation and daily editing by the author	ReRites (n.d.)

Source: compiled by the author

The final methodological section addressed ethical considerations. This stage assessed risks of bias in training corpora, potential threats to the confidentiality of archival and personal materials, and issues concerning the representation of cultural heritage. The positions of the Committee on Publication Ethics (COPE) (n.d.) and the International Committee of Medical Journal Editors (IC-MJE) (n.d.) were considered, particularly their stance that algorithms cannot be recognised as full authors and that a transparent declaration of their contribution is required.

RESULTS

Application of machine learning algorithms to the analysis of literary and historical texts

The application of topic modelling and clustering identifies latent narratives in large text corpora that remain outside the scope of conventional literary and historical interpretation. In nineteenth-century press corpora, researchers identified stable thematic patterns reflecting not only official discourses but also everyday social practices previously documented only fragmentarily (Chauhan & Shah, 2021).

Analysis of gender markers in prose fiction demonstrates long-term shifts in the representation of social roles, which correlate with broader historical changes in culture (Casadei *et al.*, 2023). Work with extensive library collections confirms the potential to trace the dynamics of genre forms and their prestige across historical periods, contributing to the development of quantitative literary studies. The use of diachronic vector spaces enables tracing changes in word meaning and usage across extended historical periods, which is particularly valuable for analysing texts characterised by evolving orthography and semantics (Miao *et al.*, 2024). Geographical reconstruction of texts shows that literary and epistolary sources contain distinct cartographic markers permitting analysis of cultural mobility and network structures. Visualisation of the routes of travellers and correspondents illustrates individual trajectories and broader spatial patterns of cultural circulation. This approach supports reconstruction of regional and transregional centres where cultural contacts concentrate and through which texts and knowledge circulate (del Rio Riande & Vitale, 2020). Spatial mapping indicates that

intertextual connections frequently overlap with geographical trajectories, forming cultural “centres of gravity” where diverse traditions intersect.

Research on repetitions and textual borrowings reveals patterns in the circulation of cultural motifs and the stability of specific textual formulations over extended periods. Analysis of Arabic-language corpora reconstructs chains of text transmission, confirming the persistence of traditions of knowledge preservation and transformation across generations (KITAB, n.d.; Romanello & Hengchen, 2021). In nineteenth-century European periodicals, identification of widespread republication describes mechanisms of cultural diffusion and the development of collective media discourses (The Viral Texts Project, n.d.). In classical Latin and Greek corpora, automatic detection of allusions documents interdependencies that remain imperceptible even to experienced philologists, expanding the interpretative potential of conventional scholarship (Riemenschneider & Frank, 2023). Analysis of manuscript sources confirms that algorithmic recognition constitutes an effective tool for producing transcriptions, although its quality depends on the condition of originals and digitisation parameters (Transkribus, n.d.). Instances of substantial errors show that automatic transcriptions without verification can distort corpus analysis. Research on OCR/HTR noise demonstrates that data quality determines the accuracy of topic modelling and authorship attribution (Liimatta, 2024). Analysis of the ECCO

collection indicates that textual distortions can alter both topic distributions and attribution results, necessitating systematic verification.

The application of algorithms in interdisciplinary projects confirms their capacity to uncover new aspects of cultural history. In the Richmond Daily Dispatch corpus, hidden narratives of home-front life during the US Civil War were reconstructed, including themes of labour mobilisation, resource scarcity, and disciplinary practices that had remained outside conventional historiography (Mining the Dispatch, n.d.; Longstreet, 2024). Studies of nineteenth-century newspaper republications reveal mechanisms of media circulation explaining the development of national cultural codes (The Viral Texts Project, n.d.). In the Arabic KITAB and OpenITI corpora, researchers demonstrated the persistence of text transmission traditions across historical periods (Open Islamicate Texts Initiative, n.d.). In classical Latin and Greek studies, intertextual analysis tools confirm that digital algorithms allow detailed mapping of literary allusions, providing empirical support for classical philology (Tesserae, n.d.). Spatial analysis using cartographic tools demonstrates the relationship between textual and geographic networks, reflecting a more comprehensive picture of cultural circulation (Pelagios Network, n.d.; GeoNames, n.d.). For summarising results and improving the visualisation of the impact of algorithmic methods on the analysis of literary and historical corpora, the material is presented in Table 2.

Table 2. Summary of the results of applying digital methods

Direction	Identified results	Cultural and interpretative effects
Thematic modelling and clustering	Latent topics in press and literature, gender patterns, genre dynamics	Reconstruction of social discourses, changes in cultural perceptions, analysis of genre prestige (Chauhan & Shah, 2021; Casadei <i>et al.</i> , 2023; Miao <i>et al.</i> , 2024)
Geographical reconstruction	Routes of travellers and correspondents, mapping textual movements	Reconstruction of cultural networks, centres of cultural circulation (Pelagios Network, n.d.; del Rio Riande & Vitale, 2020)
Intertextual analysis	Detection of borrowings, repetitions, allusions, networked connections	Understanding mechanisms of knowledge transmission and cultural diffusion (The Viral Texts Project, n.d.; Romanello & Hengchen, 2021; Riemenschneider & Frank, 2023)
Work with manuscripts	Automated transcriptions, impact of OCR/HTR noise	Confirmation of the need for corpus verification, critical engagement with data (Transkribus, n.d.; Liimatta, 2024)
Practical applications	Recovery of hidden narratives, media processes, inter-epochal connections	New approaches to the interpretation of history and literature (Mining the Dispatch, n.d.; Open Islamicate Texts Initiative, n.d.; Longstreet, 2024)

Source: compiled by the author

Contemporary digital technologies demonstrate several advantages and limitations in research application (Fig. 1). Advantages include speed and scale: platforms such as HTRC allow investigation of tens of thousands of texts simultaneously without requiring dedicated infrastructure (HathiTrust, n.d.; Project Update..., 2018). Comprehensive analysis is also significant: the combination of topic modelling, transformers for Named Entity Recognition (NER) and classification, and textual comparison algorithms produces multi-layered analysis corroborated from multiple perspectives. Novel opportunities arise from diachronic embeddings, which track changes in word semantics over time and enhance the

interpretation of long-term cultural processes. Limitations persist, notably the heterogeneity of corpora and digitisation quality, which directly affect outcomes. Low optical character recognition (OCR) quality, caused by archaic orthography, scanning defects, or errors in automatic transcription, distorts topic distributions and reduces the accuracy of stylometric analyses and authorship attribution, making preliminary cleaning and corpus normalisation critically important. Another challenge is language specificity: even the most advanced NER models, trained on modern corpora, lose accuracy on historical texts. The HIPE (Identifying Historical People, Places, and other Entities, n.d.) series demonstrated that specially

annotated data and adaptation of metrics to historical variation are required. Long documents present further limitations: basic transformers have restricted context windows, necessitating fragmentation of texts, which can result in loss of discursive dependencies (Mohammed & Ali, 2021). Interpretation of results remains challenging:

LDA topics or stylometric clusters require external verification and independent sampling. For clarity, the results are summarised in a SWOT analysis (Table 3), which allows strengths and weaknesses of digital methods to be compared with potential applications and risks affecting the interpretation of humanities data.

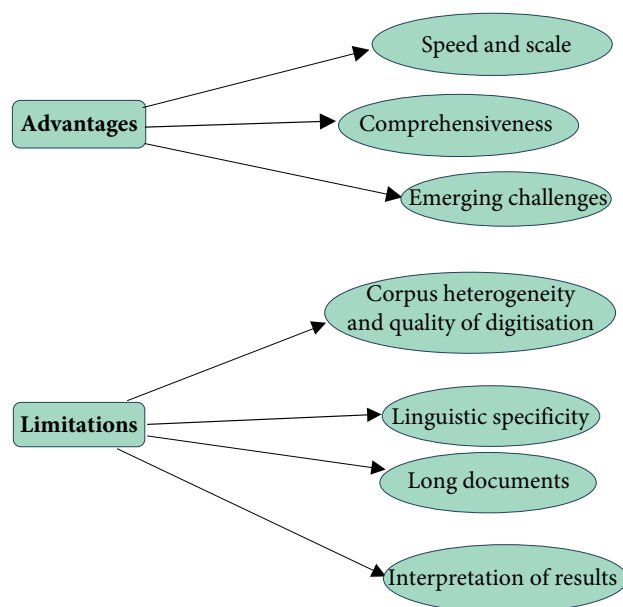


Figure 1. Key opportunities and challenges of applying digital technologies in humanities research

Source: developed by the author

Table 3. SWOT analysis of digital technologies in humanities research

Strengths	Weaknesses	Opportunities	Threats
Speed and scale of corpus analysis (HTRC allows work with thousands of texts without proprietary infrastructure)	Heterogeneity of corpora and low quality of digitisation, which distort results	Diachronic embeddings, which allow tracking semantic changes over time	Data distortion due to OCR/HTR noise, affecting result reliability
Methodological comprehensiveness (combination of thematic modelling, transformers, stylometry)	Dependence on linguistic specificity: loss of accuracy with historical texts	Expansion of cultural process analysis through integration of geographical and intertextual methods	Limited context of base transformers, complicating work with long texts
Possibility of multi-level validation of results (different methods corroborate findings)	Requirement for large-scale corpus cleaning and normalisation	In-depth study of cultural contacts using geographical reconstruction tools (Recogito)	Risk of formal interpretations of clusters and topics without external verification

Source: compiled by the author

Strengths of these technologies include scale and comprehensiveness, opening new perspectives for the study of cultural processes. Opportunities include diachronic methods that trace transformations of meanings and cultural codes over time. Weaknesses stem from low-quality source data, multilingual challenges, and the substantial effort required for corpus normalisation. Key threats include the risk of distorted interpretation due to OCR/HTR noise, limitations of transformer models with long documents, and the potential for formalist approaches to clusters and topics without further verification. This analysis clarifies the limits of digital tools while

outlining directions for further methodological development in the humanities.

Thus, the combination of topic modelling, transformers, and textual comparison algorithms opens new approaches to the analysis of large historical and literary corpora. Platforms and tools such as HTRC and Recogito standardise workflows and make digital methods accessible to a broad range of researchers. Nevertheless, digitisation quality, lack of annotated corpora for under-resourced languages, need for long-context models, and requirements for interpretative validation remain key challenges. Effective strategies include preliminary corpus cleaning and

normalisation, use of domain-adapted models, verification of results against independent data, and transparency in annotation procedures. These measures ensure reproducibility of results and pave the way for more thorough and precise research in the humanities.

Authorship and creativity in the digital age

Discussion of authorship in the context of generative systems arises from the capacity of models to produce texts that readers frequently perceive as human-created. Research demonstrates that non-expert audiences do not consistently distinguish between human and machine-generated poetry; in a large sample, participants identified the origin of poems with approximately fifty per cent accuracy, and assessments of rhythm and aesthetic quality for machine-generated texts were sometimes higher, complicating intuitive recognition of authorship (Porter & Machery, 2024). This observation directly prompts a reconsideration of conventional criteria for attributing texts to specific authors and a reassessment of the role of formal features as indicators of human labour. Normative frameworks in publishing and copyright practice establish a minimum requirement of human contribution. A policy statement and guidance from the United States Copyright Office clearly state that registration applies only to portions of a work with human creative input; sections generated by an algorithm without editorial creativity are not protected and must be declared during submission (Statement of Policy of the United States Copyright Office, 2023). Updated materials from the copyright office and public documents, including the continuation of the Report of the United States Copyright Office (2025), confirm the consistency of this approach, and an appellate decision concerning an image entirely created by a system without human input upheld the position that copyright cannot be obtained without human authorship. Collectively, these documents establish legal consensus: a machine system cannot independently hold copyright and therefore cannot constitute an author.

Editorial standards in academic publishing align with this logic. Committee on Publication Ethics (n.d.) explicitly states that language model tools cannot be listed as authors; responsibility for content and the capacity to assume obligations are incompatible with the nature of such systems (COPE Council, 2023). International Committee of Medical Journal Editors (n.d.) updated guidance requiring transparent declaration of the use of generative tools, prohibiting their listing as co-authors, and emphasising full human responsibility for verification of accuracy and mitigation of errors and plagiarism (Recommendations for the Conduct..., 2025). In the humanities, this implies clear delineation: technical assistance from a model must be disclosed in relevant sections, but the author remains the human who formulates the prompt, selects, edits, and interprets the output.

In practice, a co-creativity model emerges, viewing generative systems as partners in content creation, but not as autonomous agents. Review and empirical studies

propose conceptual frameworks that grade algorithmic involvement from instrumental assistance to collaborative creation, consistently assigning the central role to human oversight: task formulation, iterative selection of outputs, and stylistic editing (Liu, 2025). Empirical evidence from co-writing studies indicates that model involvement can enhance productivity during idea generation and structuring stages, but requires clearly defined roles and procedures to ensure attribution (Haase & Pokutta, 2024). A canonical example of co-creative poetics is the ReRites (n.d.) project, in which machine-generated material was edited daily by the author; the output was published as a twelve-book series of essays on processuality and the distribution of roles between human and model. Authorial materials emphasise rhythmic discipline, rewriting, and reprogramming of authorship through human editing of machine-generated text (Erslev, 2023). The project demonstrates that without intensive human curation, the machine's "raw" output does not attain aesthetic completion. For electronic literature studies, this case defines parameters of a hybrid genre: texts in which the algorithm's contribution is documented, but the final form depends on editorial work.

Another notable case is the novella "Death of an Author" (Aidan Marchine, 2023), produced with substantial generative tool involvement. Publisher materials and author interviews describe the process as a multi-stage assembly: models generated textual variants, while the human acted as scenario planner, editor, and compiler. The "rules of the game" method (quotas of machine-generated text, constraints on genre and composition) renders the experiment methodologically reproducible: the boundary between machine contribution and authorial selection can be traced through the protocol (Buleu, 2024). In copyright terms, the work remains human-authored due to the described editing, despite the machine origin of most of the text; this illustrates the distinction between generation and authorship. In literary and cultural studies, original experiments in poetry and drama employ models as sources of variation and combinatorial potential. Synthesised conclusions from poetry research indicate that ease of rhythm and readability metrics may positively influence non-expert perception of algorithmic texts, but such advantages do not guarantee interpretative depth; aesthetic evaluation criteria should therefore be clearly differentiated according to audience and analytical objectives (Porter & Machery, 2024). Overestimating the immutability of distinctions between human and machine texts is methodologically risky: empirical criteria should rely on independent test sets and prior disclosure of editorial procedures.

The shift towards co-creativity also affects creativity. Researchers move from a model of a single authorial act to understanding creativity as a process, in which iterative cycles of "task formulation – generation – selection – interpretation" are distributed between human and model. Procedural phases described in the literature, such as idea generation, elaboration, and realisation, formalise points where algorithmic prompts are useful, and where

exclusively human competence is required – stylistic integration, intertextual anchoring, adherence to genre and historiographical norms (Wan *et al.*, 2024). This delineation prevents erroneous attribution of creativity to the model as a capacity for responsibility; it concerns the statistical generation of variants, which only becomes a work through human selection and editorial intervention. New genres at the intersection of literature, media art, and “history-making” use generative tools as stages in a research or artistic workflow. In literature, these are hybrid collections where initial machine narratives undergo rigorous editing; in digital history, experiments with “narrative drafts” are treated as auxiliary material for further processing, rather than as final academic products. Professional communities of historians formulate guidelines for teaching and research, emphasising transparency of procedures and separation of automated generation from scholarly conclusions (American Historical Association, 2025). This reduces the risks of erroneous attribution and enhances reliability. Concerning “synthetic historiography”, it remains crucial to maintain methodological barriers between machine drafts and scholarly narratives. Special issues and research projects in digital history underscore the need to verify the instrumental role of models: from prior inventory of plots to semantic annotation of sources, with mandatory quality control, prompt documentation, and validation against independent corpora (Jones, 2023). In practice, this translates into usage protocols: modular accounting of machine-generated suggestions, versioning of iterations, and clear references to stages where human intervention is decisive.

Returning to authorship, legal and ethical consensus does not oppose intensive use of generative tools, but establishes three requirements. First, transparency: disclosure of model use in submissions or methods. Second, responsibility: final verification for factual accuracy, absence of plagiarism, and compliance with citation and copyright standards. Third, attribution: models cannot be listed as co-authors, and technical contribution must be

distinguished from creative editing. This applies to academic publications and literary editions; in the latter, paratexts such as afterwords or technical notes are required to document workflow and parameters (COPE Council, 2023). In applied practice, co-writing experiments demonstrate effects on novelty and usefulness of ideas during text preparation, but also reveal risks of stylistic homogenisation: without strict editorial frameworks, outputs may become overly predictable (Wan *et al.*, 2024). This supports the use of models for general stages – brainstorming, variant formulation, note synthesis – followed by human integration for style, rhetoric, and intertextuality.

Creativity can be distinguished at three levels (Table 4). The creative potential of models functions as a multi-level process. At the first level, they perform an auxiliary role, offering diverse linguistic variations that expand the field for subsequent selection. The next level concerns the organisation of material: algorithms assist in constructing logical flows, suggesting possible structural solutions and transitions. The highest level involves semantic work, where the defining tasks and interpretations are formulated by the researcher, alongside alignment with academic norms and source material (Haase & Pokutta, 2024). Only the combination of these levels justifies speaking of co-creation, while independent generation without authorial selection remains raw material, regardless of perception by non-expert audiences. For cultural and historical research, the co-creation perspective entails institutionalising transparent practices. Professional associations of historians propose pedagogical and research guidelines: focus on methodological compliance, documentation of automation boundaries, and clear communication to students and readers where technical assistance ends, and interpretative work begins (Jones, 2023). These principles serve as a safeguard against the mechanical transfer of machine-generated texts into scholarly conclusions while encouraging careful use of models in corpus preparation, indexing, preliminary summarisation, and heuristic analysis.

Table 4. Authorship and cultural representation in the era of AI

Level of creativity	Characterisation	Role of AI	Role of human	Examples of application	Potential risks
Operational	Provides variation in formulations, thematic moves, lexical choices	Generates alternative formulations, stylistic options, lexical combinations	Selects the most appropriate variants, verifies accuracy and contextual fit	Text rephrasing, creation of alternative titles or thesis formulations	Overload of options; risk of uncritical copying of machine suggestions
Procedural	Supports iterative structuring of text or research	Proposes plan, structure, transitions between sections, possible presentation schemes	Determines final logic of presentation, edits structure according to purpose and standards	Article plan generation, options for argumentative transitions, assistance in building logical blocks	Formal correctness without deep alignment to topic; risk of losing originality of authorial intent
Conceptual	Shapes the content framework and semantic substance of research	Can suggest associative ideas or alternative interpretations	Defines cognitive frameworks, works with sources, ensures compliance with disciplinary and methodological standards	Choice of research paradigm, definition of theoretical approach, interpretation of results	Distortion of content through superficial generation; risk of substituting scientific interpretation with machine simulation

Source: developed based on COPE Council (2023); C.R. Buleu (2024)

In the age of generative systems, authorship is less about “who wrote the text?” and more about describing the production process: who formulated the task, which tools were applied, where editing occurred, and which parts were verified and validated by humans. Legal and editorial positions (guidance from the United States Copyright Office, COPE, and ICMJE policies) standardise the requirement for substantial human contribution and transparency. Research and artistic cases, such as ReRites and Death of an Author, serve as sites for clarifying the boundaries of co-creativity. For the humanities, this means generative models can support stages of idea creation, variant formulation, and initial drafting, but co-author status remains with the human who defines the concept, assumes responsibility, and ensures interpretative validity.

Ethical challenges in the use of AI in cultural research

The use of AI systems in cultural research generates several ethical challenges, encompassing bias, privacy, and cultural representation. The primary issue is bias. Training data typically originate from archives, museum and library collections, often created within specific historical periods dominated by certain cultural narratives. This reproduces the same imbalances in automated systems. Bias can be present in cultural heritage collections and their digital replicas, and the AI pipeline can amplify these distortions (Foka & Griffin, 2024). Researchers should apply bias mitigation techniques at each stage of the process: from data collection to annotation and modelling. Research on practical experience with AI systems in the cultural sector highlights the challenge of the “black box”, which limits transparency (Gîrbacia, 2024). This complicates the assessment of why a model produces a given result and how it may reproduce structural biases present in source data. In this context, the concept of the cultural semiosphere by J. Lotman (Bankov, 2020) is relevant, emphasising that each culture functions as an integrated semiotic system with its own centres and peripheries. Use of corpora formed around dominant cultural narratives means AI models reproduce and reinforce these dominants, marginalising peripheral voices. This renders bias not only a technical issue, but also a semiotic-cultural one, as automated algorithms participate in the redistribution of meaning within the semiosphere.

Another aspect concerns privacy and data protection. Archival and personal texts, which may contain personal data or be subject to copyright restrictions, are used to train models. In the context of cultural heritage, this raises questions of accessibility and transparency, since some systems operate as closed platforms without explanations of who can access the data or how it is used (Yang *et al.*, 2023). This relates to Z. Bauman’s idea of liquid modernity (Doyle & Conboy, 2020), according to which social structures and rules constantly change, and stable forms of control are replaced by flexible but less transparent mechanisms. The opacity of algorithmic platforms reproduces this fluidity, creating situations in which conventional cultural memory institutions, such as libraries and archives, lose clear

contours and fall under the influence of new, less predictable forms of data governance. Studies addressing intellectual property in the United Kingdom emphasise that model training often relies on copyright-protected texts, which can lead to uncertain legal circumstances: who has access, on what basis the data are used, and whether models have legal rights to train on archival texts (Westenberger & Farmaki, 2025). The result is corrections that create uneven cultural representation and affect research reliability.

A third set of challenges concerns ethical dilemmas in cultural representation. The use of artificial intelligence in the reconstruction or reproduction of cultural heritage can distort authenticity. This occurs due to insufficient understanding of local contexts, misinterpretation of symbols or images by algorithms (Floridi, 2023). More broadly, this aligns with cultural flows articulated by A. Appadurai (Linares, 2025), which describe the movement of ideas, images, and symbols across global cultural spaces. Algorithms involved in heritage reconstruction become instruments of these flows, but risk extracting artefacts from their local contexts and transforming them into universal, yet simplified, cultural signs. Consequently, the use of algorithmic models may transform authentic local meanings into globalised formats, heightening ethical risks in representation. It is necessary to establish boundaries of responsibility, preserve historical integrity, and involve cultural experts in interpretative processes. In terms of academic integrity, there are risks to the originality of texts: when AI generates works, there is a temptation not to declare its contribution, which violates professional standards. Furthermore, lack of transparency, for example, over who generated or edited material, or whether the model had any significant contribution, can lead to plagiarism or incorrect attribution (Gallent-Torres *et al.*, 2023). This creates the need for protocols requiring transparent documentation of machine contributions as part of authorship or research processes.

Ethical application of machine learning systems in cultural research therefore requires formalisation of protocols, including minimising bias through quality control of corpora, annotation, and model customisation that accounts for cultural diversity; data protection and transparency, including labelling confidential materials and documenting components of generative models to ensure reliability; preservation of authenticity, separating machine reconstruction from historical context and involving cultural experts in verification; and attribution and integrity, with explicit declaration of AI roles, human responsibility for results, and thorough documentation of processes. Implementation of such protocols is achievable through interdisciplinary collaboration, where experts in the humanities, ethics, law, and technology jointly develop principles that combine methodological rigour with historical and cultural sensitivity. This ensures not only the legality and correctness of AI applications but also ethical accountability towards cultural heritage, audiences, and future researchers.

DISCUSSION

The results indicate that combining topic modelling, transformer-based approaches to NER/EL (Named Entity Recognition/Entity Linking), stylometry, and tools for identifying textual correspondences produces a coherent picture of genre-thematic configurations, networks of textual borrowing, and variations in authorial signal, while delineating ethical boundaries for AI use in the interpretation of cultural data. The study demonstrates the convergence of multiple methodological lines – LDA/clustering, long-context models, stylo analysis, and passim comparison – in reaching consistent conclusions through independent verification and shows reproducibility across different corpora, including newspapers, Arabic texts, classical corpora, and manuscript transcripts. Analysis of authorship and creativity further indicates that generative models function effectively as co-creation tools at operational and procedural levels, while ultimate significance rests with the human, who ensures conceptual integrity, interpretation, and accountability. Findings concerning ethical challenges emphasise that bias in corpora, privacy risks, and the danger of distorted cultural representation require systematic protocols for labelling, transparency, and expert involvement.

With respect to topic modelling, the results confirm that LDA and related clustering methods are suitable for the stable induction of latent topics in large corpora and for mapping discourse evolution. Comparison with the approach of I.C. Chang *et al.* (2021) shows methodological proximity: their study employed text mining, clustering, and LDA for thematic classification of journals in environmental education, reflecting the practice of using LDA for initial inductive segmentation of the corpus. Similarities include the three-step procedure of preprocessing, clustering, and LDA interpretation, along with the validation of topics using domain-relevant indicators. Differences lie in the corpus domain: I.C. Chang *et al.* examined journals with clearly defined sections, whereas the present study considered newspaper and historiographical corpora, which exhibit greater genre and lexical variability and require diachronic embeddings and more stringent normalisation prior to topic modelling. In authorship attribution, stylometric results confirm the utility of most frequent words, Δ -metrics, and sliding-window stylometry for localising shifts in authorial signal within a single work. Comparison with H. Ilseman (2023) shows alignment in tools, including the use of the stylo package for R and attention to attribution challenges in complex corpora. Similarities include the evidential basis of stylometric decisions and the need for transparent parameters. Differences concern material and task framing: H. Ilseman addressed the English-language canon and the authorship question of “Henry V”, where textual traditions and editorial layers are well documented, whereas the present study addresses multilingual data, including newspaper OCR/HTR texts and Arabic corpora, which require additional quality filters and post-OCR cleaning prior to stylometry. Passim

analysis complements stylo, linking attributional conclusions with networks of repeated use.

Analysis of human + AI co-authorship shows that increased productivity in the ideation stage can coincide with risks of style unification, aligning with findings by K. Medeiros *et al.* (2025), who empirically evaluated creativity dynamics with ChatGPT. Similarities include recognition of novelty and utility gains at the generation and structuring stage, alongside the need for human integration to maintain stylistic coherence. Differences appear in evaluation type: K. Medeiros *et al.* operationalised measures through experimental design, whereas the present study relies on corpus metrics, including topic/cluster consistency and stability of stylometric profiles. Another difference concerns the purpose of co-authorship: in the present study, it supports tasks such as drafting analytical paragraphs and producing annotation variants, whereas K. Medeiros *et al.* (2025) examined AI-assisted creation in the context of general creative activity. From an authorship theory perspective, normative conclusions regarding the primacy of human contribution and the need for paratextual transparency correspond with T. Kamatović (2024), who discussed authorship ambivalence in the AI era through engagement with the Augustinian tradition. Similarities include recognition of the shift from a monolithic authorial “I” to a procedural schema, where decisions about meaning and interpretation are made by humans. Differences concern argumentative grounds: the present study emphasises empirical and legal dimensions, including responsibility, attribution, and the non-representativity of AI as author, whereas T. Kamatović draws on posthumanist frameworks and retrospective theological-philosophical analysis.

In literary theory, A. Gefen (2024) examined the transformation of the author’s role under algorithmic mediation. Alignment with the present study occurs in the reduction of the authorial centre and the growing importance of editorial, curatorial, and verification practices, consistent with a procedural co-creation model. Differences concern the scale of generalisation: A. Gefen emphasised cultural-potological diagnosis, whereas the present study links findings to specific methodological techniques and attribution protocols. Conceptual conclusions about shifts in authorship gain empirical support from data showing the division of the process into operational, procedural, and conceptual levels. Regarding cultural representation, the section on ethical risks in reproducing cultural narratives with generative tools resonates with C. Oztabak (2025), who analysed AI-generated folklore narratives and their cultural relevance. Similarities include recognition of potential for cultural resonance when careful curatorial frameworks and local expertise are present. Differences concern emphasis: the present study prioritises labelling algorithmic reconstructions as interpretations and requires source verification, whereas C. Oztabak focused on bridging tradition and digital storytelling with an applied orientation. This highlights the distinction between institutional scientific validation and project-based practices of cultural engineering.

Concerning linguistic and cultural mediation, the results on shifts in semantic fields and the necessity of human interpretation for cross-linguistic transformations align with B. Wei (2024) on the strengths and weaknesses of AI-mediated creative translation. Similarities include recognition of algorithmic usefulness in variant paraphrasing and stylistic suggestion. Differences concern accuracy requirements: the present study prioritises disciplinary control, including adherence to terminology, source material, and historiographical norms, whereas B. Wei emphasises creative advantages and general technical challenges, such as style preservation and cultural nuance, with less focus on rigorous academic validation. In the domain of algorithmic bias in multimodal systems, observations on cultural distortions align with W. Elsharif *et al.* (2025), who summarised strategies for identifying, evaluating, and mitigating bias in text-to-image models. Similarities include recognition of the systemic nature of cultural skew and the conclusion that the full cycle – from data curation through metrics to interventions – determines outcomes. Differences concern modality: the present results are text-centred, whereas Elsharif *et al.* examine the visual aspect. Both approaches converge on the need for clear protocols for transparency, labelling, and expert involvement to ensure accurate cultural representation.

In summary, integrating algorithmic methods into humanities research proves productive under three conditions: methodological triangulation, infrastructural support, and ethical regulation, including transparency, attribution, and participation of domain experts. Importance lies in demonstrating that technical tools enhance heuristic capacity, while final academic meaning emerges through human interpretation. Alignment with other studies is evident in conclusions concerning the utility of inductive topic methods, the capabilities of stylometry, the constructive yet limited effect of co-authorship, and the delineation of ethical boundaries in cultural representation. Differences stem from corpus domain, data modality, and validation criteria.

CONCLUSIONS

The study systematically analysed the potential of machine learning algorithms for the analysis of literary and historical texts and identified problematic aspects related to authorship, creativity, and ethical challenges. The obtained findings confirm that combining topic modelling, transformer-based models, stylometric methods, and tools for identifying intertextual links allows not only the processing

of large corpora but also the generation of qualitatively new insights into genre dynamics, cultural circulation processes, and transformations of authorial practices. The study established that the use of topic modelling methods, particularly LDA, reliably identifies latent topics in large corpora, traces their temporal dynamics, and recovers hidden narratives.

It was confirmed that neural models based on transformers, including BERT, Longformer, and BigBird, serve as effective instruments for text classification, named entity recognition, and analysis of discursive features. These models achieve accuracy in identifying toponyms, persons, and organisations, and in classifying genre markers. Full engagement with historical corpora, however, requires specially annotated data that account for orthographic and syntactic variability. Results also indicated that infrastructural solutions, such as the HathiTrust Research Center, hold strategic importance for humanities studies.

Concerning authorship and creativity, the study established that generative systems can function at stages of variant formulation, planning, and draft creation, while final conceptual integrity, interpretation, and responsibility remain human. Co-authorship is conceptualised at three levels: operational, concerning variation in formulation; procedural, concerning structuring and support of iterative processes; and conceptual, concerning intention and interpretation. The study demonstrated that generative models constitute effective instruments, but cannot be regarded as autonomous authors. Ethical analysis identifies bias in training corpora, confidentiality risks, and challenges of cultural representation as genuine threats to research using AI.

The study was limited by its reliance on the synthesis of existing academic sources and case examples. Ethical concerns related to confidentiality and representation could not always be fully addressed within the scope of the study. Future studies should expand practices using long-context models, develop infrastructural solutions, and deepen ethical reflection to ensure greater accuracy, transparency, and cultural relevance in research of this kind.

ACKNOWLEDGEMENTS

None.

FUNDING

None.

CONFLICT OF INTEREST

None.

REFERENCES

- [1] American Historical Association. (2025). *Guiding principles for artificial intelligence in history education*. Washington: American Historical Association.
- [2] Bankov, K. (2020). Platfospheres and sociocultural explosion of Web 2.0: The commercial centre of the digital semiosphere. *Sign Systems Studies*, 48(2-4), 246-270. doi: 10.12697/SSS.2020.48.2-4.04.
- [3] Buleu, C.R. (2024). *Literary works generated by artificial intelligence: Methodology, achievements, dilemmas*. *Dacoromania Litteraria*, 11(1), 90-107.

- [4] Casadei, P., Bloom, M., Camerani, R., Masucci, M., Siepel, J., & Ospina, J.V. (2023). Mapping the state of the art of creative cluster research: A bibliometric and thematic analysis. *European Planning Studies*, 31(12), 2531-2551. [doi: 10.1080/09654313.2022.2158722](https://doi.org/10.1080/09654313.2022.2158722).
- [5] Chang, I.C., Yu, T.K., Chang, Y.J., & Yu, T.Y. (2021). Applying text mining, clustering analysis, and latent dirichlet allocation techniques for topic classification of environmental education journals. *Sustainability*, 13(19), article number 10856. [doi: 10.3390/su131910856](https://doi.org/10.3390/su131910856).
- [6] Chauhan, U., & Shah, A. (2021). Topic modeling using latent Dirichlet allocation: A survey. *ACM Computing Surveys*, 54(7), 1-35. [doi: 10.1145/3462478](https://doi.org/10.1145/3462478).
- [7] Chun, J., & Elkins, K. (2023). The crisis of artificial intelligence: A new digital humanities curriculum for human-centred AI. *International Journal of Humanities and Arts Computing*, 17(2), 147-167. [doi: 10.3366/ijhac.2023.0310](https://doi.org/10.3366/ijhac.2023.0310).
- [8] Colella, S. (2025). The language of the digital air: AI-generated literature and the performance of authorship. *Humanities*, 14(8), 1-17. [doi: 10.3390/h14080164](https://doi.org/10.3390/h14080164).
- [9] Committee on Publication Ethics (COPE). (n.d.). Retrieved from <https://publicationethics.org/>.
- [10] COPE Council. (2023). *Authorship and AI tools*. [doi: 10.24318/cCVRZBms](https://doi.org/10.24318/cCVRZBms).
- [11] del Rio Riande, G., & Vitale, V. (2020). Recogito-in-a-box: From annotation to digital edition. *Modern Languages Open*, 1, article number 44. [doi: 10.3828/ml.o.v0i0.299](https://doi.org/10.3828/ml.o.v0i0.299).
- [12] Doyle, R., & Conboy, K. (2020). [Considering the merits of applying 'liquid modernity' as a theoretical lens for information systems research](https://doi.org/10.3390/ijerph18052488). In *Proceedings of the 28th European conference on information systems (ECIS 2020), research-in-progress papers* (Paper 90). Marrakech: Association for Information Systems.
- [13] Duch, W. (2024). [Artificial intelligence and the limits of the humanities](https://doi.org/10.3390/ijerph18052488). *Er(r)go. Teoria-Literatura-Kultura*, 48, 269-297.
- [14] Elsharif, W., Alzubaidi, M., & Agus, M. (2025). Cultural Bias in text-to-image models: A systematic review of bias identification, evaluation and mitigation strategies. *IEEE Access*, 13, 122636-122659. [doi: 10.1109/ACCESS.2025.3585745](https://doi.org/10.1109/ACCESS.2025.3585745).
- [15] Erslev, M.S. (2023). [A poetics of misrepresentation](https://doi.org/10.3390/ijerph18052488). In A. Ensslin, J. Round & B. Thomas (Eds.), *The Routledge companion to literary media* (pp. 197-208). London: Routledge.
- [16] Floridi, L. (2023). *The ethics of artificial intelligence: Principles, challenges, and opportunities*. Oxford: Oxford University Press.
- [17] Foka, A., & Griffin, G. (2024). AI, cultural heritage, and bias: Some key queries that arise from the use of GenAI. *Heritage*, 7(11), 6125-6136. [doi: 10.3390/heritage7110287](https://doi.org/10.3390/heritage7110287).
- [18] Frontoni, E., Paolanti, M., Migliorelli, L., Pietrini, R., & Asimakopoulos, S. (2024). Artificial intelligence: The new frontier in digital humanities. *Frontiers in Computer Science*, 6, article number 1529826. [doi: 10.3389/fcomp.2024.1529826](https://doi.org/10.3389/fcomp.2024.1529826).
- [19] Gallent-Torres, C., Zapata-González, A., & Ortego-Hernando, J.L. (2023). The impact of generative artificial intelligence in higher education: A focus on ethics and academic integrity. *Electronic Journal of Educational Research and Evaluation*, 29(2). [doi: 10.30827/relieve.v29i2.29134](https://doi.org/10.30827/relieve.v29i2.29134).
- [20] Gefen, A. (2024). The author's second death. *Poetics Today*, 45(2), 301-307. [doi: 10.1215/03335372-11092938](https://doi.org/10.1215/03335372-11092938).
- [21] GeoNames. (n.d.). Retrieved from <https://www.geonames.org/>.
- [22] Gîrbacia, F. (2024). An analysis of research trends for using artificial intelligence in cultural heritage. *Electronics*, 13(18), article number 3738. [doi: 10.3390/electronics13183738](https://doi.org/10.3390/electronics13183738).
- [23] Haase, J., & Pokutta, S. (2024). Human-AI co-creativity: Exploring synergies across levels of creative collaboration. *ArXiv Preprint*. [doi: 10.48550/arXiv.2411.12527](https://doi.org/10.48550/arXiv.2411.12527).
- [24] Hannaford, E.D., Schlegel, V., Lewis, R., Ramsden, S., Bunn, J., Moore, J., & Nenadic, G. (2024). Our heritage, our stories: Developing AI tools to link and support community-generated digital cultural heritage. *Journal of Documentation*, 80(5), 1133-1147. [doi: 10.1108/JD-03-2024-0057](https://doi.org/10.1108/JD-03-2024-0057).
- [25] HathiTrust Research Center Analytics. (n.d.). Retrieved from <https://analytics.hathitrust.org/>.
- [26] HathiTrust. (n.d.). Retrieved from <https://www.hathitrust.org/>.
- [27] HIPE – Identifying Historical People, Places and other Entities. (n.d.). Retrieved from <https://surl.li/eyguhe>.
- [28] Ilseman, H. (2023). R Stylo and the authorship determination of Henry V. *Digital Scholarship in the Humanities*, 38(4), 1577-1581. [doi: 10.1093/llc/fqad067](https://doi.org/10.1093/llc/fqad067).
- [29] International Committee of Medical Journal Editors (ICMJE). (n.d.). Retrieved from <https://www.icmje.org/>.
- [30] Jones, M.L. (2023). AI in history. *The American Historical Review*, 128(3), 1360-1367. [doi: 10.1093/ahr/rhad361](https://doi.org/10.1093/ahr/rhad361).
- [31] Kamatović, T. (2024). Augustine and the posthuman writer: Toward an ambiguous authorship in the age of artificial intelligence. *Journal of Posthuman Studies*, 8(1), 6-22. [doi: 10.5325/jpoststud.8.1.0006](https://doi.org/10.5325/jpoststud.8.1.0006).
- [32] KITAB. (n.d.). Retrieved from <https://kitab-project.org/>.
- [33] Lavidas, K., Voulgari, I., Papadakis, S., Athanassopoulos, S., Anastasiou, A., Filippidi, A., & Karacapilidis, N. (2024). Determinants of humanities and social sciences students' intentions to use artificial intelligence applications for academic purposes. *Information*, 15(6), article number 314. [doi: 10.3390/info15060314](https://doi.org/10.3390/info15060314).

- [34] Liimatta, A. (2024). [OCR quality and the resilience of algorithmic identification of linguistic register features in nineteenth century collections online](#). *Journal of Data Mining & Digital Humanities*.
- [35] Linares, L. (2025). Bridging disjunctures: Appadurai's cultural anthropology of globalisation and literary translation studies. *Translation in Society*, 4(1), 97-113. doi: [10.1075/tris.24016.lin](#).
- [36] Liu, Z. (2025). Human-AI co-creation: A framework for collaborative design in intelligent systems. *ArXiv Preprint*. doi: [10.48550/arXiv.2507.17774](#).
- [37] Longstreet, J. (2024). *From Manassas to Appomattox: Memoirs of the civil war in America*. Bloomington: Indiana University Press.
- [38] Marchine, A. (2023). *Death of an author*. New York: Pushkin Industries.
- [39] Medeiros, K., Cropley, D.H., Marrone, R.L., & Reiter-Palmon, R. (2025). Human-AI co-creativity: Does ChatGPT make us more creative? *The Journal of Creative Behavior*, 59(2), article number e70022. doi: [10.1002/jocb.70022](#).
- [40] Merriam, T. (2021). [Review of Stylo\(\). Other, by M. Eder, J. Rybicki, & M. Kestemont](#). *Renaissance and Reformation*, 44(4), 223-228.
- [41] Miao, X., Zhu, S., Fu, F., Guo, Z., Yang, Z., Tu, Y., & Cui, B. (2024). [X-former elucidator: Reviving efficient attention for long context language modeling](#). In *Proceedings of the thirty-third international joint conference on artificial intelligence* (pp. 8179-8187). Bremen: International Joint Conferences on Artificial Intelligence.
- [42] Mining the Dispatch. (n.d.). Retrieved from <https://dsl.richmond.edu/dispatch/>.
- [43] Mohammed, A.H., & Ali, A.H. (2021). [Survey of bert \(bidirectional encoder representation transformer\) types](#). *Journal of Physics: Conference Series*, 1963(1), article number 012173.
- [44] Open Islamicate Texts Initiative. (n.d.). Retrieved from <https://openiti.org/>.
- [45] Oztabak, C. (2025). [AI-generated folklore for culturally resonant storytelling in the digital age: Bridging tradition and technology in storytelling](#). In E. Saka (Ed.), *Understanding generative AI in a cultural context: Artificial myths and human realities* (pp. 161-184). New York: IGI Global Scientific Publishing.
- [46] Pelagios Network. (n.d.). Retrieved from <https://pelagios.org/>.
- [47] Porter, B., & Machery, E. (2024). AI-generated poetry is indistinguishable from human-written poetry and is rated more favorably. *Scientific Reports*, 14, article number 26133. doi: [10.1038/s41598-024-76900-1](#).
- [48] Project Update – Bentham vs the computer. (2018). Retrieved from <https://surl.li/yiqrmd>.
- [49] Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. (2025). Retrieved from <https://www.icmje.org/icmje-recommendations.pdf>.
- [50] Report of the United States Copyright Office "Copyright and Artificial Intelligence: Part 2: Copyrightability". (2025, January). Retrieved from <https://surl.li/fnctzm>.
- [51] ReRites. (n.d.). Retrieved from <https://www.glia.ca/rerites/>.
- [52] Riemenschneider, F., & Frank, A. (2023). Graecia capta ferum victorem cepit. Detecting latin allusions to ancient greek literature. *ArXiv Preprint*. doi: [10.48550/arXiv.2308.12008](#).
- [53] Romanello, M., & Hengchen, S. (2021). Detecting text reuse with passim. *Programming Historian*. doi: [10.46430/phen0092](#).
- [54] Shahid, A., Ogunola, A.A., Ishaq, S.M., Khawaji, T., Mowafaq, F., & Arafat, Y. (2024). [Revolutionizing education: The transformative power of artificial intelligence in shaping a brighter future for humanities](#). *Library of Progress-Library Science, Information Technology & Computer*, 44(3), 21898-21912.
- [55] Statement of Policy of the United States Copyright Office "Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence". (2023, March). Retrieved from <https://surl.li/qevfxv>.
- [56] Tesserae. (n.d.). Retrieved from <https://www.tesserae.co/>.
- [57] The Viral Texts Project. (n.d.). Retrieved from <https://viraltexts.org/>.
- [58] Transkribus. (n.d.). Retrieved from <https://www.transkribus.org/>.
- [59] Wan, Q., Hu, S., Zhang, Y., Wang, P., Wen, B., & Lu, Z. (2024). "It felt like having a second mind": Investigating human-AI co-creativity in prewriting with large language models. *Proceedings of the ACM on Human-Computer Interaction*, 8, 1-26. doi: [10.1145/3637361](#).
- [60] Wei, B. (2024). [Advancements and challenges in AI-driven creative translation: A comprehensive analysis](#). *Applied and Computational Engineering*, 82, 82-87.
- [61] Westenberger, P., & Farmaki, D. (2025). Artificial intelligence for cultural heritage research: The challenges in UK copyright law and policy. *European Journal of Cultural Management and Policy*, 15, article number 14009. doi: [10.3389/ejcmp.2025.14009](#).
- [62] Yang, J., Zhang, X., Liang, K., & Liu, Y. (2023). Exploring the application of large language models in detecting and protecting personally identifiable information in archival data: A comprehensive study. In *2023 IEEE International Conference on Big Data (BigData)* (pp. 2116-2123). Sorrento: IEEE. doi: [10.1109/BigData59044.2023.10386949](#).
- [63] Zhao, H. (2023). Challenges or opportunities: When artificial intelligence is applied to digital humanities. *Journal for ABM Special Issue*, 8(1), 57-65. doi: [10.33063/tabm.v8i1.548](#).